DOCUMENT RESUME

ED 340 777 TM 018 033

AUTHOR Mitchell, Ruth; Stempel, Amy
TITLE Six Case-Studies of Performance Assessment.
SPONS AGENCY Congress of the U.S., Washington, D.C. Office of Technology Assessment.
PUB DATE Mar 91
NOTE 227p.; Contractor report prepared for the Office of Technology Assessment titled "Testing in American Schools: Asking the Right Questions." For related document, see TM 018 025.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC10 Plus Postage.
DESCRIPTORS *Case Studies; Comparative Analysis; Cost Effectiveness; *Educational Assessment; Elementary Secondary Education; Minority Groups; *Portfolios (Background Materials); *Program Evaluation; Psychometrics; *School Districts; Scoring; Social Studies; Student Evaluation; Teacher Attitudes; United States History
IDENTIFIERS Advanced Placement Examinations (CEEB); Arizona; Authentic Assessment; California Assessment Program; *Performance Based Evaluation; San Diego Unified School District CA; South Brunswick Township School District NJ

ABSTRACT

Six cases are studied to give some idea of the range of performance assessments in use in the United States. Contrast and comparison are made possible through the study of the following: (1) the South Brunswick (New Jersey) Public Schools Observational Portfolio for kindergarten through grade 12; (2) the Arizona Student Assessment Program (not yet operational); (3) the California Assessment Program as it relates to the San Diego City Schools; (4) the College Board's Advanced Placement (AP) Studio Art Portfolio Assessment; (5) the Longfellow Elementary School Art Assessment in Bozeman (Montana); and (6) the "We the People..." Competition, a national social studies-history and oral discourse performance assessment organized by the Center for Civic Education (CCE) in Calabasas (California). Each study contains: (1) an introduction; (2) a statement of purpose and rationale; (3) the experience of program implementation; (4) a description of the impacts on teachers; (5) comments on scoring; (6) cost issues; (7) a discussion of impacts on minorities; and (8) a discussion of the lessons from each case study. Two general lessons are that the purpose of the assessment must be clear before decisions are made about form, and that the variety of performance assessments causes many psychometric complications. A 57-item bibliography is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340222

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by
Ruth Mitchell Ph.D. and Amy Stempel
Council for Basic Education

for
Office of Technology Assessment
U.S. Congress

March 1991

TM018033

This contractor document was prepared for the OTA assessment entitled *Testing in American Schools: Asking the Right Questions*. It is being made available because it contains much useful information beyond that used in the OTA report. However, it is not endorsed by OTA, nor has it been reviewed by the Technology Assessment Board. References to it should cite the contractor, not OTA, as the author.

BEST COPY AVAILABLE

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

INTRODUCTION

Assessment of student achievement and program quality in American schools is in a state of upheaval. What formerly seemed the business of test publishers and administrators has in the past two or three years become everyone's business. Orthodoxies are being challenged and assumptions questioned. A new vocabulary has to be acquired for a discourse in a different arena and among different speakers.

The upheaval had its origins in a collision between demands for accountability made by legislatures who monitor educational expenditure and the simultaneous development of curricula promoting conceptual understanding and application, not memorization. These parallel developments can be dated from the middle 1980s, when the two strands emerged as responses to the National Commission on Excellence in Education's report, A Nation at Risk, published in April 1983. The educational system was being asked to show numerical results and at the same time produce thinking, creative students. The assessment instruments available seemed inadequate to the task. For example, the Arizona Department of Education found that the norm-referenced multiple-choice tests it was using for accountability measured only 30%, at best, of its revised curriculum framework.

Out of a complicated turmoil of developments, including the dissemination of the whole-language approach to literacy; the publication of statements on curricula by professional organizations such as the National Council of Teachers of Mathematics and the American Association for the Advancement of Science; a reconsideration of testing young children; and the results of international evaluations unflattering to American education, a movement towards naturalistic evaluation emerged. The first (and still the most widespread) example was the direct assessment of writing. Writing samples began to replace multiple-choice tests, which were perceived as tests of editorial knowledge.

At first the movement called for "alternative" assessments, for obvious reasons, and then the search for an accurate descriptor led to "authentic." Neither these adjectives nor the awkward "performance-based assessment" hit the mark. We prefer the expression "performance assessment" because of its accuracy and its resonance with real-world speech: we speak of a car's or an economy's "performance." Nevertheless, the expression lacks zip, as we are the first to acknowledge, and exactly what it covers is always in dispute. For example, some writers include portfolio

assessments as a subset of performance assessments, and for others, they are a separate category.

There is however a central area of agreement on performance assessments, despite fuzziness at the edges: performance assessments, unlike multiple-choice, can take many forms, but they are all designed to assess directly what students know and are able to do. If students should be able to analyze, synthesize, interpret, and evaluate facts and ideas, then they should be asked to demonstrate these abilities. If they should be able to solve problems using mathematical knowledge, then the assessment must require problemsolving. Performance assessments include direct writing; conversations in a foreign language; observations, for example of students conducting science experiments, or of children's litereacy development; open-ended questions, where students provide solutions for a problem which has no single right answer; group projects where students show they can work cooperatively; interviews, giving students a chance to reflect on their achievement; portfolios, selections of student work; and other less easily categorized assessment tasks.

Like all forms of educational evaluation, performance assessments are intended to provide information. The major purposes for which information is needed in educational settings are: program evaluation; feedback on instructional strategies; and student selection and sorting, often called "certification." As our case-studies will show, performance assessments can and do fulfill each of these purposes. But they may also have another purpose, the modeling of desired curriculum and instruction. In this they differ radically from norm-referenced multiple-choice examinations. "Teaching to the test" is regarded as poor pedagogical practice when the tests are multiple-choice, but it is encouraged when they have been designed specifically to encourage teachers to adopt teaching centered on concepts and applications.

Performance assessments are spreading rapidly at all levels of education. A recent survey by the Council of Chief State School Officers revealed that 39 states now have some form of performance assessment at the state level, usually a writing sample; 15 states have performance assessments in other subjects besides writing; five states, California, New York, Kentucky,, Maryland, and Arizona have declared a policy of gradually moving to performance assessments for state accountability. Connecticut has a consortium with 20 other states and the Coalition for Essential Schools to develop performance assessments for high school mathematics and science, which will be adopted by Connecticut for accountability when their design is complete. Vermont has adopted portfolio assessment of writing and mathematics for state accountability. These numbers do not include states which are thinking about performance assessments, nor the districts and individual schools which have devised their own assessments, using their own personnel or cooperating with test publishers.

Thus, when we chose these six examples of performance assessment for study, there was already a pool to draw from. By the time the studies are published, there may be as many as ten times more examples.

Since performance assessment is a technology in an early stage of development, there is no coherent systematic distribution of examples--so many at the elementary, so many at the secondary level, so many in English/language arts, so many in mathematics, for example. Various different forms of performance assessment arise to meet needs which may be political or social as much as educational. They may depend on personalities or apparently fortuitous opportunities. The present ad hoc picture may soon be organized at least at the state level by a network to be administered by the Council for Chief State School Officers, but it will still be difficult to keep track of developments at the district and school level.

To choose a typical sample for study seemed impossible and becomes harder as the number of

examples increases. We also wanted to add to the collection of models available. Enough information seemed to be in the public domain about the Vermont statewide portfolio assessment; the Connecticut high school mathematics and science group assessment tasks; the Rockefeller Foundation funded ARTS PROPEL portfolio project, undertaken jointly by the Educational Testing Service (ETS), Project Zero at Harvard Graduate School of Education, and the Pittsburgh Public Schools; the Rites of Passage Experience (ROPE) at Walden III Alternative High School in Racine, Wisconsin; and the cumulative performance assessment at Alverno College (although not a K-12 assessment, this has exerted a strong influence on K-12 programs).

The cases we chose give some idea of the range of performance assessments being developed, but do not claim typicality. We will explain our choices in an order intended to provide context by heightening contrasts and comparisons.

We begin with the South Brunswick Public Schools Observational Portfolio for Grades K-2. This is a district program to assess young children using developmentally appropriate methods. We wanted to know how it worked, how teachers used observation to assess their students without drowning in paperwork, and how the school community viewed the portfolio as an assessment tool.

We next move to a statewide program, the Arizona Student Assessment Program (ASAP), which is not operational yet but has the specific purpose of influencing teaching by providing models of instruction in the form of assessments. We wanted to see how a state program was presented to teachers and also to understand how a state department of education and a test publisher cooperate.

In contrast, we chose to look not at the California Assessment Program (CAP) as such, but at its political relationship with a large powerful and progressive school district, San Diego City Schools (SDCS). The CAP/SDCS study focusses on the interaction of a state assessment program with a city school district as they both struggle to develop an integrated system of assessment to meet their specific needs. We wanted to know how they exert pressure on each other towards performance assessment and what obstacles they each encounter.

From the state and district level, we move to the national level and an example of a longstanding successful national performance assessment--the College Board's Advanced Placement (AP) Studio Art Portfolio Assessment. Developed and administered by ETS (like all the Advanced Placement examinations), AP Studio Art has no multiple-choice and no written portion. A team of college and high school faculty evaluates the contents of the portfolio and arrives at a single score for the portfolio as a whole. We wanted to know how it works and what effect the program has on teachers and students.

A different aspect of art, learning how to appreciate its nature, is the subject of the Longfellow Elementary School Art Assessment in Bozeman Montana. A single elementary school has devised a method of assessing knowledge about art by using parents as interviewers. While not polished or sophisticated psychometrically, the assessment illustrates a creative solution to a local problem.

Our final assessment, the "We the People..." Competition organized by the Center for Civic Education (CCE) in Calabasas, CA is an anomaly. It is a national competition funded by the Bicentennial Commission, with stages of competition at the Congressional District, state, and national level. Both the curriculum supplied by CCE and preparation for the competition take place in schools, but they are not part of the regular school program. There are a number of such competitions--the Odyssey of the Mind, the Physics Olympiad, History Day, and the Westinghouse Science Competition--which all use imaginative performance tasks and which seem as if they would

provide excellent models for both teaching and assessment. We wanted to see how the "We the People..." Competition works, since it involves community members as judges, and to assess its potential as a measure of both history-social studies learning and oral discourse which might be incorporated into the formal educational system.

We gathered information from site visits; material developed to assist the implementation of the assessment (sample tasks, instructions, forms, etc.); and telephone conversations. Because performance assessments are in their infancy, there is almost no formally published material describing and analyzing them. This is especially true for the six cases we chose for this study. The bibliography in the final appendix provides general background material, useful for an introduction to the topic in general, but directly related only to one or two of these cases.

We have arranged them in a format with as much uniformity as their variety would permit. Because of this variety (deliberately sought, as we said above), some of the sections will contain little information in specific cases. Each study begins with an introduction, explaining its nature and what features drew us to focus on it. Section II looks at the purpose of the assessment and the rationale for the choice of a particular form. As might be expected, the purpose dictates the other features of the assessment as we focus on them in turn. For example, the Section III describes the experience of implementing the assessment, which flows logically from the purpose which it serves. In this section the reader will find the major factual information about the program studied.

Section IV discusses the impact on teachers, again deriving from the purpose. Section V approaches the issue of scoring, which, with the exception of the Advanced Placement Studio Art Portfolio Assessment, has received somewhat less attention than the design of assessment tasks, and therefore presents unresolved problems.

Section VI attempts to elucidate the issue of costs, a major concern to those contemplating moving towards performance assessments. Unfortunately it is almost impossible to compare costs either with other forms of assessment or across our case-studies, for two major reasons: the different purposes of the assessments mean that the money to pay for them comes from different parts of educational budgets; and categories are not comparable--what one school district or state department of education includes as the costs of assessment is not counted as such by another. We have raised questions and pointed directions for further research which may make some order out of baffling and contradictory data.

Section VII discusses impacts on minority populations and possible legal issues, although in most of our cases the performance assessments are in such preliminary stages that challenges have not yet arisen. Section VIII concludes with a discussion of the lessons to be learned from the case-study.

A summary and analysis, also following the same division into sections, attempts to derive some generalizations from the case-studies. It looks at what the case studies tell us about the success of performance assessments, the problems they present, and the factors affecting the solution of the problems.

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

I. SOUTH BRUNSWICK PUBLIC SCHOOLS:

A CASE STUDY OF OBSERVATIONAL ASSESSMENT

Section I. Introduction

We chose the South Brunswick Public Schools, New Jersey, elementary school assessment for study because it represents assessment designed to complement developmentally appropriate instruction for young children. It uses adults' observations to evaluate the child's progress in language acquisition, emphasizing growth over time rather than single-point testing.

There is a growing class of these observational, multi-faceted records of early childhood learning, (usually K-3). The states of Georgia, North Carolina, and Missouri have statewide early-childhood assessments, similar to each other but with local variations, and six school districts in Alaska are experimenting with forms of these assessments, with the assistance of the Alaska Department of Education.

The Primary Language Record

To a large extent, these statewide assessments have been developed by groups of teachers (led by state department of education personnel) responding to the mismatch between developmentally appropriate instruction and multiple-choice, machine-scorable assessments for young children.

At the same time, there is a parallel development: the introduction of a model from England, called the Primary Language Record (PLR), designed at the Center for Language In Primary Education (CLPE) in London. The California Literature Project has endorsed the PLR and is offering a series of workshops for its participants led by the designers of the PLR, Myra Barrs and Ann Thomas. The California Literature Project is a statewide professional development program based on the Bay Area (now National) Writing Project model. The California Literature Project is working to replace basal readers with children's literature, and embraces the whole-language approach to literacy development. For these reasons, the PLR seems particularly appropriate as a record of progress. Individual school districts and schools on the East Coast, such as the Central Park East schools, are also consulting with Barrs and Thomas to bring the PLR into their literacy programs.

The attraction of the PLR is a systematic method of organizing the observations teachers make routinely. The PLR consists of two parts, a continuous working record and a summary form, completed three times a year. The working record includes observations of the child's literacy

behavior, such as "running records" of reading aloud, and writing samples, as well as a list of books the child can read, either in English or a home language. The summary record includes an interview with the parents about what the child likes to read and do at home, and an interview with the child about interests. The interviews take place at the beginning and end of each school year. The summary record goes with the child to the next grade, throughout primary school, ages five through 11.

In London, the PLR is used in schools to report the child's progress to parents and to maintain a profile for the school's records. There are no multiple-choice assessments corresponding to the norm-referenced tests used in U.S. elementary schools, although some tests may be used at age 11 to determine placement in the comprehensive secondary schools. There has been a recent attempt to introduce a national examination system in England, including an assessment at age 7. Since this is in the experimental stage, it is possible that the PLR and its expansion into mathematics and science--the Primary Learning Record--could influence the design of the early childhood part of the national examination, and even replace it.

South Brunswick Public Schools did not adopt the Primary Language Record as such, but used it as a basis for its observation portfolio. Parts of the portfolio are directly taken from the Primary Language Record, but are used in the context of an instrument designed by South Brunswick teachers for their own circumstances.

South Brunswick Public Schools

South Brunswick Township is a municipal school district located in New Jersey halfway between New York and Philadelphia, a geographical location which influences the economic base of the township. The town depends on high-tech industries in the area and the services required to maintain those industries. Princeton University and Rutgers University (the state university of New Jersey) are both within easy commuting distance.

On the whole there is neither extreme poverty nor extreme wealth, although a wide socioeconomic spectrum is represented. There are large houses with swimming pools which sell for half a million dollars, but mobile homes are barely concealed by the luxuriant suburban foliage. Ethnically, the district is mostly white (73%), with 10% Black, and 17% Asian (a large number of them Indians, attracted by the technological industries). In the entire district there are 60 ESL students who speak 36 different languages.

The Township covers 41 square miles and contains six K-6 elementary schools, one middle school, and one high school. District-wide enrollment is 3800; 2300 of the children are in the elementary schools. South Brunswick is a suburban area composed of what used to be many small towns which grew into each other as a result of the economic growth of the 1960's. Many people commute to the two universities, some to New York City,
and some remain in the area in a string of industries based on microchip technology.

South Brunswick Public Schools does not attempt to replace multiple-choice tests with its portfolio. The district maintains the tests currently mandated by the State of New Jersey. All students in every grade must be comparatively tested in order to determine which students require compensatory education and which schools will receive extra funds for those students. The South Brunswick Public Schools administers Circus Math and How Much? How Many?, tests developed by the Educational Testing Service, the Concepts of Print test, and a writing sample, in kindergarten; the new Stanford Test in the first grade; and the CTBS4 in the second grade in mathematics and language arts. These tests are norm-referenced, multiple-choice tests used nationwide.

The fact that South Brunswick Public Schools keeps two assessment systems in place (the portfolio and the published multiple-choice tests) caused us to ask about the comparative value of each kind of assessment to teachers, administrators, and parents. But because the focus of this study is performance assessment, it will concentrate primarily on the assessments being conducted in South Brunswick Public Schools and how those assessments relate to the multiple-choice testing within the district, not on the district's testing policy and practice in general.

Section II. Purpose of and Rationale for the Assessment

The purpose of the South Brunswick Public Schools portfolio is to show the progress of language acquisition in young students K-2. The portfolio is used to track progress and to focus instruction on those areas where individual students are having trouble. There is no intent to select and sort students using this portfolio, nor is there any attempt to evaluate the program directly, although there are constant adjustments within the classrooms using it.

The portfolio is not scored yet, but plans are being formed in conjunction with Educational Testing Service researchers to produce a numerical literacy scale based on the portfolio.

The form of the portfolio and its contents have evolved from exploring methods of accurately reporting the literacy development of young children. There may have been some original hope that the Primary Language Record would work for South Brunswick, but the teachers modified it from the start. The rationale for an observational assessment emerged from the history of South Brunswick's developmentally appropriate curriculum, which will be discussed in the next section.

Section III. South Brunswick Public Schools' Experience to Date with the Portfolio

The move towards performance assessment originated with Willa Spicer, director of curriculum and instruction. In the early eighties she charged the teachers of South Brunswick Public Schools with researching and implementing a more developmentally appropriate early childhood education curriculum.

Developmentally appropriate curriculum views the young child (before grade 3) as a natural learner whose methods of learning differ from those of older children--and from one another. Young children learn by doing and performing, have a low tolerance for boredom, and live in the literal world. The assumption behind developmentally appropriate teaching is that the more methods of teaching are tied into the developmental stages of children the more effective those methods will be in imparting knowledge and skills to children.

The result in South Brunswick Public Schools was a move to this type of curriculum and teaching in the lower grades (K-2). (The South Brunswick Public Schools' developmentally appropriate curriculum gained national attention when it was featured in a Newsweek article on children's learning, 17 April 1989.)

In language arts, developmentally appropriate instruction implies the "whole language" teaching method. "Whole language" is an eclectic approach to developing literacy, using the child's own language and experiences, instructional strategies such as phonics and word recognition where necessary, substituting children's literature for basal readers, and focussing on meaning. In whole

language classrooms, children do not fill out worksheets identifying vowels and consonants, and are not given spelling lists to memorize. South Brunswick Public Schools early childhood programs embrace the whole language approach, and are piloting a new textbook by Rigby which is a "whole language" approach to learning mathematics.

Spicer realized that a new method of teaching required a new method of assessment consistent with developmentally appropriate instruction. Since all testing exercises an influence on what is taught and how, there was concern that developmentally appropriate teaching could be thwarted by an emphasis on the discrete knowledge and skills required to succeed on machine-scorable, multiple-choice tests. Knowing that she could not evade the tests which the state of New Jersey requires, she wanted nevertheless to reduce their importance as curricular influences by shifting attention to growth over time in children's literacy.

She consulted with Dr. Edward Chittenden, senior researcher at the Educational Testing Service in Princeton, who introduced her to the Primary Language Record. Chittenden has remained close to the South Brunswick Public Schools project as the teachers developed instruments during summer workshops, field-tested them in their classrooms, and then modified them. He developed the Word Awareness Writing Activity (WAWA) with the teachers, giving them a means of documenting children's growth in connecting the sounds of words with their appearance in written form.

In all six elementary schools, K and 1 are participating in the portfolio assessment as a matter of district policy. Grade 2 will be required to incorporate the portfolio into the classroom beginning in 1991.

At present, and until the report card revisions are in place, the portfolio is used to adjust instruction for each child, to conference with parents and other adults who work with the child, and to assess the child's growth over time rather than to assign a grade to the student. In essence, the portfolio is an instrument which allows individualized instruction and the measurement of student growth over time as opposed to a single-point evaluation.

The portfolio was used for the first time in 1989 with kindergarten teachers. Its introduction is cumulative, with a grade added each year. Its use is mandatory, but there are supports in the form of workshops to reduce the threat and discomfort.

The portfolio caused a good deal of distress in the school year 1989-90 because the paperwork load seemed overwhelming. But, as one teacher said to us, "South Brunswick truly is a teaching district," and Spicer responded to complaints by scheduling a summer workshop to revise the portfolio before it was used in first grade. By chance, there was a representative from all six schools at the summer workshop, so that there was no sense of railroading unwilling schools. Revisions were proposed and tried until both kindergarten teachers who had used the portfolio and the grade 1 teachers who were scheduled to use it were satisfied that it was workable.

<u>The South Brunswick Public Schools K-2 Observation Portfolio</u>

South Brunswick's assessment system is referred to by everyone concerned as "portfolio assessment," although it consists of much more than the term usually brings to mind (a collection of student work in one subject, on the whole). Although student work is found in the South Brunswick Public Schools portfolio it is primarily a teacher portfolio. The student's teacher makes the decisions about what the portfolio contains and administers the portfolio evaluations to the students. A student portfolio on the other hand, is created from student works usually chosen by the student for representation in the portfolio.

The portfolio's origins are eclectic: it is heavily influenced by the Primary Language Record but has additional elements, one specially developed by Chittenden. Like the statewide K-3 assessments in North Carolina, Missouri, and the Alaskan districts, the South Brunswick Public Schools portfolio was developed primarily by teachers, who looked at the Primary Language Record and Reading Recovery (a project of Ohio State University), but then designed what they needed. It has already undergone revisions in a 1990 summer workshop, and will undoubtedly be modified again.

The South Brunswick Public Schools portfolio does not include student's self-evaluation of selected works to be included. If the portfolio extends through the upper grades (Spicer mentioned this in passing with the caution that it would be far in the future) then students would gradually begin to take more responsibility for their portfolio development as they progress through the grades.

The school district hopes to have a mathematics portfolio in place by the end of the 91-92 academic year and by that time also to begin a Best Work Portfolio, in which students keep their highest accomplishment in all subject areas. The mathematics portfolio is being modeled on the principles of the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics.

South Brunswick Public Schools does not have a uniform district-wide report card format for grades 1-12, so each school is revising its first grade report cards this year to reflect the portfolio's use in the classroom.

There are nine parts to the portfolio. (Examples of all these materials are in the appendix) :

1. Self-portrait. The child is asked to draw a self-portrait, which is then affixed to the portfolio (or even drawn directly on the manila folder). The exercise is repeated at the end of the year. There is no formal evaluation, of course; the exercise provides an objective record of how the child's self-concept and eye-hand coordination changes over a year. Teachers are able to effectively assess the student's level of development by looking at the self-portrait. One kindergarten teacher said, "If a child is attaching his arms to his head, then you know he's not ready to visualize concepts about print."

2. Child Interview. A teacher or other adult in the classroom (at Deans School a high school student in the CIPED program was interviewing the children) writes down answers the child gives in response to questions about favorite activities, TV watching, reading, and the literacy activities of others in the home.

3. Parent questionnaire. This is sent home for parents to fill out before their first conference with the teacher. It asks parents about their child's reading interests so the parent may develop a working relationship with the child's teacher.

4. Concepts About Print. This brief test is administered by reading two specially adapted books with the child. It assesses the child's concepts of print before learning to read. Children are asked to point to the front and back of the book, to show where the story begins, to indicate the direction in which print is decoded, and to explain the meaning of question marks and periods (if sufficiently advanced). The book includes pages printed upside down and with the print running in the reverse direction.
   The test has a dual function: it is part of the portfolios, and it is part of the mandated package of tests for kindergarten children in South Brunswick Public Schools which is used

to identify children in need of compensatory education (it is also a nationally normed test in New Zealand).

5. The Word Awareness Writing Activity (WAWA) records the level at which children begin to comprehend the rules of forming words. The teacher asks the children to write a word after hearing it repeated and used in sentences. They are reassured that they are not expected to know the spellings but are expected only to try to render in writing what they hear. Children's progress is recorded in five stages: precommunicative--random letters or scribbling; semi-phonetic--some sounds represented by letters; phonetic--letters used appropriately for sounds; transitional--some evidence of awareness of spelling patterns; mostly correct--10 out of 13 words correct. This test was developed with Chittenden, in order to demonstrate that there is a natural progress towards acquiring understanding of what written language represents (sounds), so teachers and parents should not be alarmed at scribbles and invented spelling.

6. Reading Sample. This one-to-one reading assessment allows the teacher to estimate the strategies each individual child is using to deal with print. The teacher is required to take a reading sample and document it with analysis three times a year, but can administer additional assessments if the child seems to be having difficulties. There are two forms the teacher may choose from when administering the Reading Sample: the Running Record or Miscue Analysis.
The Running Record is used with emergent readers, students who mimic the act of reading but do not yet know how to read. Running Records can be administered on a known book or an unknown book depending on the information the teacher wants. Using a known book reveals to the teacher how well the child makes use of reading strategies (what to do with an unknown word, how to use the context to infer meaning), and is used with children in the very early stages of reading. Using an unknown book reveals the child's ability to integrate reading strategies independently and shows a movement toward independent reading. The Running Record provides a means of recording what a young child is thinking and doing while reading.
The Miscue Analysis is used with independent readers, that is, students who are self-correcting and reading fluently.
For a Miscue Analysis, the book must be unknown to the student. The teacher records the miscues using an accepted code, and then analyzes them for clues to the student's instructional needs.
One form for recording a child's reading growth is photocopied directly from the English Primary Language Record.

7. Writing Sample. This is a sample of a student's free writing. It is unedited by the teacher, but the teacher may provide a written "translation" if invented spelling and sketchy syntax make it difficult to read easily.
One form for recording a child's writing growth is photocopied directly from the English Primary Language Record.

8. Class Record. This is the only element in the portfolio which is not a part of the child's individual record. It is a class profile enabling the teacher to identify those children who may need extra attention in certain areas. The record is a one-page matrix, with yes-no answers to five questions: does the student pay attention in large and small groups? interact in groups? retell a story? choose to read? write willingly? The teachers use copies of the Class Record at regular intervals to adjust their class instruction distinguished from the

individual attention they give each student.

9. Student Observation Forms. On the basis of information obtained from the class record, teachers choose observation forms which will enhance their knowledge of individual student discrepancies. Teachers are not required to use these forms and usually only do so when they need more information about a student's development.

Physically, the portfolio is a legal-sized accordion-pleated folder in dark red, filled with manila folders marked with the names of each element. There is one portfolio for each teacher, and an individual one for each child, filled with manila folders for the collection of materials, and copies of interviews and tests. Teachers keep the whole collection near their desks: "I just run to it when there's a question about a child--it's right there in a milk crate," one teacher told us.

Each teacher's portfolio also contains optional observation sheets which are to be attached to the reading and writing samples if the teacher thinks more observation is necessary. Whenever a child is having difficulty, the teacher is expected to keep detailed records of the child's activities, thus making it less likely that a child will be lost in the system.

The portfolio provides teacher, parents, and school administrators with both individual and group data on the literacy development of their students. Students are observed interacting in the classroom, problem-solving, reading aloud, and writing. Numerical grading is not used, and the impact of mandated multiple-choice testing is reduced to a few sessions of familiarization with number 2 pencils and "bubbles" before the tests are given.

Section IV. Impacts on Teachers and Instruction

We visited four elementary schools, each of which gave us a different piece of information about the place of the portfolio assessment in the schools and how it affects the teachers and instruction.

Deans Elementary School provided us with a picture of a school unified in purpose and intent upon the implementation of the portfolio by the second grade. At Brunswick Acres Elementary School on the other hand, there was some hesitation about the merits of the portfolio and some unevenness in its application. At Constable Elementary School we were able to see the administration of some of the components of the portfolio to individual students. Greenbrook Elementary School provided us with parents willing to discuss their admiration and concerns for the new portfolios.

The following discussion uses the individual schools as a framework, but freely moves between them and the whole district. The discussion of our interviews at Deans will seem disproportionate when compared to the treatment of the other schools, where we spent an equal amount of time--more, in the case of Brunswick Acres, where we attended Back to School Night. What we learned at Deans should be read as general background for the South Brunswick Public Schools elementary schools. We received a full exposition of the portfolio and its use at Deans, and nothing we heard or saw later contradicted it, except where noted.

Deans Elementary School

At Deans Elementary School we first met with teachers Roberta Duffy and Joan Warren, who enthusiastically described the development and the effects of the developmentally appropriate curriculum and the portfolio. Deans, a school with 170 students, has a new principal Patricia Holliday, who thoroughly supports the assessment. Under the former principal, the staff made a proposal for the use of compensatory education funds, and consequently

has a teacher's aide or paraprofessional full-time in each classroom.

Two important themes emerged from our discussion with Duffy and Warren, themes found throughout our experiences in South Brunswick. One was the teachers' participation in all aspects of curriculum and portfolio development, including the freedom not to use textbooks if a teacher prefers not to (although the portfolio is mandated); and the other was the teachers' definition of assessment as part of teaching.

On the first theme, the Deans teachers were eloquent about the teachers' ownership of their professional tools. "I love South Brunswick," said Duffy, who is in third year of teaching there after experience in other school districts. "They don't hand out portfolios and say 'do them.' We have workshops all the time to show how to do them." (One of the workshop leaders is a teacher at Constable School, whom we later watched administer running records and miscue analyses.) The portfolio was originally developed in summer 1989 by groups of teachers, with the assistance of Chittenden of the Educational Testing Service and the constant encouragement of Spicer. It was radically revised in 1990, when the interviews and the questionnaire for the parents were added, and a writing exercise called "Write all the words you know," was deleted.

In South Brunswick, "teachers feel listened to and supported," said Duffy. Teachers in the six elementary schools meet informally together once a month to exchange problems and solutions they find in using new textbooks and materials. The line between compensated time and volunteer attendance is apparently blurred, but our two teachers were not troubled by this.

Warren added that when the early childhood program was being set up in the early 1980's, small groups of teachers would experiment with different materials and approaches. "It reminded me of a Bank Street setting," she said. "We would 'play' with the materials, and find that everyone chose the same one without a mandate from Spicer." At the same time, there is freedom to use whatever materials the teacher is comfortable with--some teachers still use basal readers.

On the second theme, it was apparent that assessment to these South Brunswick Public Schools teachers is part of teaching. In fact it was difficult to focus them on assessment as accountability. The teachers do not centralize grading, even for report cards. "I don't think it's for grading, but for me to teach them what they need to know. And I know what they need, because it's all there in the portfolio," said Duffy. The portfolio provides "tangible proof of what you're talking about," the observations made in the classroom. She uses it for referrals to the Student Assistance Team, a child study group which includes a psychologist, a social worker, and a counselor; conferences with parents; to improve her own teaching; to know exactly the weaknesses and strengths of the children; and to identify children who need special attention such as the Reading Discovery program (South Brunswick's version of Reading Recovery).

The teachers thought that national testing was not useful to them because it does not assess development in whole language. "The tests go from part to whole, and our programs go from whole to part," Duffy summarized. "Those tests are basically for basals, and to assess these kids that have learned whole language by basals--it makes no sense at all."

But they regarded the mandated tests as irrelevant more than as obstacles. They realize that parents accept the multiple-choice scores as the authoritative evaluation of their children, but they believe that the portfolios could become replacements for multiple-choice tests, if parents understood how they function. In fact, from conversations and interviews with parents, (with one notable exception), the function seems clear to them and there may be hints of change in the parents'

blanket acceptance of multiple-choice tests. South Brunswick Public Schools itself is certainly moving towards making the portfolio a formal assessment instrument in its research with Chittenden and the Educational Testing Service, a move which should reassure parents.

When we saw it being used, the portfolio was in effect a mechanism to drive individualized instruction, not an assessment instrument. It provides a record of student progress to be used from day one of the school year to day 180 and even across years. Using it, teachers find it easy to put their finger on the learning strategies of each child and consequently to develop activities which will stress the student's strengths while providing practice and help with weaknesses. In the past, teachers have taught and assessed primarily verbal and logico-mathematical learning, but the portfolio allows them to understand and support other kinds of learning styles.

The portfolios induce a different approach to retention issues: when the time comes for retention decisions, the portfolio record is consulted to see if the child has made progress. If progress can be shown, then the student is promoted, on the assumption that every child develops at his own rate and there is really nothing to worry about until the child reaches the third grade. If no progress is apparent, the child is promoted but is identified for compensatory education.

The emphasis on literacy development goes beyond the portfolio. Each child in grade 1 and up at Deans writes in a journal every morning, an activity which the two teachers identified as another assessment giving them information they need to help the children learn.

Observations in the classrooms of Deans Elementary School revealed immediately one of the reasons for the success of the observation portfolio: the presence of several adults in the classroom. This was a noticeable feature of all three schools we visited. At Deans, it was the result of a deliberate policy of using compensatory education funds to pay for paraprofessionals.

Space limitations mandated no more than 20 children in the Deans classrooms, and each room had, in addition to the teacher, either an aide, or a student teacher, or a high school student, and sometimes one each of these. One classroom had a "student grandmother," a retired woman who gives a few hours each day to assisting. The additional personnel make individual attention feasible, as well as providing a number of adults for the children to interact with.

In one class, a high school student was administering the Child Interview portion of the portfolio while the teacher worked with another group on manipulative activities. The high school student was a participant in the Community Involvement Program for Educational Development (CIPED), a program in which all juniors in South Brunswick High School spend Wednesday in community career exploration. A number of students spend time in the schools, where they work under a teacher's supervision exactly like paraprofessionals.

In the same kindergarten room, one small boy was thoroughly enjoying himself alone at the paint table, smearing paint all over his hands, making handprints on paper, and then covering the entire sheet with paint. The teacher said what she had learned about the child from using the portfolio indicated that he had not enjoyed the advantages of free play with messy, squishy materials that many of his middle-class peers had experienced, so she was letting him catch up.

One kindergarten teacher showed us how her children had performed on the Word Awareness Writing Assessment (WAWA), and how she was able to use the information from it to adapt instruction to the needs of each individual child. One child put an "h" in front of every word she heard, another could almost spell the word "monster." One child, identified for help by the Reading

Discovery team, seemed to be at a more advanced stage of literacy than another who had not been identified, so the teacher decided to discuss their cases with the team.

Brunswick Acres Elementary School

Brunswick Acres has a large three-year grant from the RJR Nabisco Next Century Schools Project to bring about changes in the school. The school will use the grant to encourage risk-taking among the teachers, who are now working in teams to brainstorm about the use of the Next Century Schools money.

The school has 370 students in a sprawling maze-like building. Some seemed to view another new thing with cynicism: some have been at the school for more than 20 years, and, in the words of a grade 1 teacher, have an attitude which says: "Here's something new again. Maybe we tried it ten years ago in a different package, and here it is, showing its face again." The grade 2 teacher said she was already trying out elements of the portfolio and thought she would be able to use it in the Fall.

Kindergarten and first grade teachers seemed enthusiastic about the portfolio, with the exception of a new teacher who was experiencing problems. A first-grade teacher showed us a "big book" written by the children in her class about South Brunswick. She said "Learning has to be fun--teaching has to be fun too." She allows children to take home big books to read, since they come in sets of four copies, impressing on them the proper care and return of borrowed books.

Somewhat to our surprise, this teacher, who has been involved in the development of the portfolio from its beginning, as both kindergarten and first-grade teacher, believed that it could and should replace multiple-choice for accountability purposes at this level. Her opinion was colored by a recent visit to her daughter's high school Back to School night, where she was dismayed the pressure of grades on the students: "it's as if I have to think about a trick to memorize the definition of this word, so that I can get a good mark on the test, instead of learning for the sake of learning. I was so frustrated, knowing how we assess kids with the portfolio at this level."

In the evening, we sat in Edie Polumbo's class and watched her explain the portfolio to her kindergarten parents. She introduced it as district policy, since it is mandated, but described it as "like a standardized test, only better." The parents did not ask many questions about it, although they seemed aware of it and approving.

Afterwards we spoke with a South Brunswick Public Schools School Board member, John Wolf, who had a daughter in Polumbo's class. He liked the portfolio concept, believing that it provided more complete information about students. But he did not think that it could ever replace norm-referenced multiple-choice tests. He did not offer any reasons, but seemed surprised that replacing multiple-choice tests should even be suggested.

Constable Elementary School

The discussion of parent understanding of the portfolio continued next morning when we met Dr. Judith Zimmerman, principal of Constable Elementary School, which has 405 children, pre-kindergarten to grade 6. She advocates continued parent education, which is a feature of the Constable program. In May of the kindergarten year, she invites parents for an extended discussion of developmentally appropriate education. Her grade 1 teachers provided workshops in reading and mathematics for the parents to show them how children learn and teach the strategies that parents can use at home to enhance their child's learning.

Nevertheless, parents are influenced by their own experience of school. They want to see report cards, which are currently being revised to incorporate the portfolio's records. One parent has been carrying on a campaign in Dr. Zimmerman's office to ensure that his daughter is given hours of homework. He recently arrived waving a William Raspberry column saying that children don't work hard enough as evidence that children need homework. Dr. Zimmerman tries to explain that a lot of homework is inappropriate at this age, but thinks the parent may send the child to a private school where he can get his way.

While this incident occurred in the upper grades, it is indicative of the tension between the old and the new. The old norm-referenced, multiple-choice tests focus on breaking knowledge down in order to measure it. The numbers are clear and understandable. The new developmentally appropriate methods of teaching and testing do not have the perceived rigor of the old tests but are systemically valid. Until this tension is eliminated, full implementation of the portfolio system will be slow.

At Constable we watched the teacher mentioned at Deans, Sharon Suskin, as she sampled the reading skills of three children in her grade 1 class of 19 children, nine of them compensatory education students. She is a reading specialist self-taught in Reading Recovery strategies, now training other teachers, in Reading Discovery, the South Brunswick Public Schools version of Reading Recovery. The Reading Recovery program (the subject of a feature article in Education Week, 7 November, 1990, pages 1 and 11), is based on the work of Marie Clay at Ohio State University. Reading Recovery is short-term intensive one-to-one attention to children having difficulty with reading, designed to return them after a specific time to the middle reading level of their class. Reading Discovery has the same aims but a shorter training period. Suskin trains her fellow teachers and district specialists in district-financed workshops.

Reading Recovery requires a year-long training program to enable teachers to diagnose and treat the reading difficulties of individual children. The two programs, Reading Recovery and Reading Discovery, will run side by side in South Brunswick Public Schools, thus making it possible to compare their effectiveness. Suskin's Reading Discovery certainly has the advantage of being economical in a straitened fiscal climate. As she puts it, "We may not achieve the results that the year-long training at NYU would ensure, but we're still doing ten thousand times better than we did before!" "What we did before" included waiting for the problem to correct itself, or retaining students until they master the material.

Suskin sat at a round table in one corner of her classroom while a full-time substitute teacher worked with the rest of the children. The school has a grant to employ a full-time substitute so that teachers can give special attention to at-risk children.

Suskin characterized her teaching as continuous assessment and response to the information it provides. All the reading (there are no reading groups) in her class is either literature or related to what the children are studying, whether social studies, science, or mathematics. She praised the Rigby Mathematics Series because it incorporates the use of language into mathematics learning.

Suskin explained how she organizes her teaching so that she can use the portfolio as an integral part of instruction.
At one point in the morning she settles the class down and then calls children one by one to her table for individual conferencing. She can see half of the class (about 10 children) every day from 10:30 to 11:50 a.m, since some children only require two minutes of her time and others require ten. She may see three children together if they are all reading the same book, to ask them to retell the story, or to respond to some events in it. It is during this time that the portfolio

assessments take place.

Suskin and the paraprofessional who assists for an hour each morning keep dated class record sheets at hand all the time to make ethnographic notes on the children as they work with them, especially the compensatory education children. About once a month, these notes are summarized for each child and put into the child's folder. They are used for two main purposes: to conference with parents, and to pull together impromptu groups of children who may all be having the same difficulty, so that they can be helped simultaneously.

As we watched, she asked each child to read and then made a running record as they did so. The children then responded to comprehension questions about what they read. One child was superficially proficient in reading, but the running record revealed that he was decoding rather than reading for meaning; that he went back to the beginning of phrases he misread rather than correcting them on the spot; and that he did not seem to be enjoying the story. Suskin made notes to talk with the child's parents about his reading experiences at home. Another child was at a much less advanced stage of reading, but was beginning to recognize words and understand that he could look at other places in the book to find the same words in a new context.
Another teacher, Daniel Diamond (the only male teacher we interviewed), has incorporated a writing portfolio for early elementary children into his instruction based on the theory and practice of Lucy Calkins and Donald Graves. This is an example of the South Brunswick Public Schools's encouragement of individual teacher initiative: the principal, Dr. Zimmerman, mentioned Diamond's portfolio to us, as did other teachers. The writing portfolio is not an assessment tool, but a means of documenting the development of children's writing behaviors. It is not germane to the case-study, except as an illustration of the district's attitude towards allowing its teachers to experiment with different pedagogies.

Greenbrook Elementary School
Here we met with four parents, who were invited by the principal, Stephanie Craib on Spicer's suggestion, to meet with us. They left their work in order to sit at a small table in a tiny closet-like room for ten or fifteen minutes of discussion with us--in itself a tribute to their support for the school.

The first couple, Philip and Diana Allsopp, were enthusiastic about both the developmentally appropriate instruction and the portfolio assessment. Their child had transferred to Greenbrook from a somewhat rigid experience in a Montessori school and found the atmosphere open and accepting. He told his parents that "I can't do anything wrong," meaning that he met only with encouragement, not criticism. The parents said that the child has grown in confidence, does not feel pressure, and is acquiring the skills that he needs--that is, the child's instruction is personalized as a result of the portfolio assessment.

Mr. Allsopp confessed to a prejudice against standardized testing because "they teach children to take tests, but I'm not sure they do anything for encouraging creativity, which is where the future of this country lies."

During a second interview, a mother, Sandy Popko, expressed the only explicit negative reaction we heard to the portfolio assessment. She wants assurance that there is standardization and that her daughter is learning what everyone else in the country is learning. She believes that traditional tests are not subjective at all and that the portfolio cannot provide any assurance of objectivity. She thinks that if teachers develop, administer, and grade tests they can do what they want and report what they want. If the family moves, her daughter may be at a disadvantage because other schools

will not understand how her daughter has been taught. She also fears that her daughter will be at a disadvantage when she takes the SAT and the other standardized, multiple-choice tests which at present determine one's future.

At the same interview a father with a doctorate in education, Ron Woods, responded to Popko that he thought it more important to teach skills than impart knowledge at the early childhood stages. He was not looking for a comparison of his daughter to others but assurance that she knows how to do certain things and how to think for herself. He likes the portfolio because it provides much more information about the child's development than standardized tests, which he regards as useless.

Section V. Scoring Issues

Since the state mandates yearly testing, the district cannot phase out multiple-choice tests. "The challenge," says Spicer, "is to not let the multiple choice test scores stand for more than they are worth." In order not to make change too threatening, the standardized test scores are recorded on reports to the district and the parents, with additional space for the results of non-standardized tests and performance assessments.

While the portfolio will continue to evolve as a feedback mechanism to improve instruction, it also has potential as an accountability instrument. Spicer and Chittenden are beginning a research project to devise a means of aggregating data from the portfolios. They are examining literacy scales, among them a scale worked out in London by Barrs and Thomas, and will rank children's literacy as evidenced by the portfolio on these scales. Then teachers in an adjoining school district which does not use portfolios will be asked (and paid) to use the scales and rank the portfolios, in order to evaluate how well the system communicates standards. Further work will establish the South Brunswick-Educational Testing Service (as Spicer and Chittenden call it) scale for evaluating children's progress in literacy.

It is unlikely that the scale will soon replace multiple-choice, norm-referenced tests for accountability in South Brunswick Public Schools, since these are mandated in New Jersey, but the existence of aggregatable data will enhance the scope and value of the portfolio.

As of now, the South Brunswick Public Schools portfolios are not scored. They are used by teachers to tailor instruction to individual students. Because of their purpose, we do not see this lack of scoring as a negative but rather a result of the system not being geared towards standardized reporting.

Standardization of the portfolio assessment was not an issue for the teachers, because of its perceived role as instructional information rather than accountability tool. There may be an assumption of implicit standardization because teachers have been involved with the initial design and subsequent revision of the portfolio. Since they agreed on what went into it and how it should be used and interpreted, it seems to be assumed that it is used in a standard fashion throughout the six schools. As far as our observations and conversations went we concluded that it is being used in a standardized way. It is not an issue that preoccupied the teachers or the administration. The parent who objected to the use of the portfolio as sole assessment was concerned with national norms rather than with standardization as such.

There are additional approaches to standardizing the portfolio. To begin with, some of the contents, the Concepts of Print test and the WAWA can be scored using a key. Then running records and

miscue analyses can also be scored without much difficulty. Those aspects of the portfolio which cannot be scored using a key--the writing sample, for instance--can be graded by a group of teachers developing a rubric from each set of papers. Thus the contents of the portfolio could be assessed using standardized scores, as well as the points on literacy scales.

There could also be standardized by moderation, that is, by exchanging a sample of portfolios among teachers, so that they each read about 10% from each class and could then discuss their common standards. This is the method used to maintain standards in a number of European school systems. The method of checking a random 10% sample was used by the New York State Education Department to ensure standardization of the results of the grade 4 science Manipulative Skills Test.

If the portfolio gains status as an accountability tool, it will be necessary to consider how to equate the portfolio's standards to national norms. (Aggregation of data is the objective of the South Brunswick Public Schools-Educational Testing Service research project.) This is an enormous question facing all advocates of performance assessment, and is at present also preoccupying the U.S. Department of Education's division concerned with the assessment of Chapter 1 students. While the answers to equating differing assessments to each other are not clear yet, it does seem that they will need to take a different form than classic psychometric methods.

Section VI. Economic Issues

The cost of the portfolio is difficult to disentangle from the South Brunswick Public Schools budget. Like most performance assessments in our case study, the costs are spread throughout the budget in different ways than the costs of tests purchased from test publishers. The following section provides a roadmap of the costs of the portfolio.

During the first summer, 10 teachers were paid for two days of time to design the portfolio, but a great deal of the work had already been done during normal school days. The cost for summer 1989 was therefore about $2100. In Summer 1990, it rose to almost $4,000 because six teachers spent most of two weeks working on the revisions. Chittenden's consulting time in both summers was paid out through a grant to Educational Testing Service from the Bruner Foundation.

Ongoing workshops which help teachers analyze reading assessments such as running records and miscue analyses are compensated with credits for future raises on the salary scale. A series of after-school workshops costs about $4,000. If the workshop instructor is an South Brunswick Public Schools employee (Sharon Suskin, for example), she is paid for both preparation and instructional time at the usual rate of $21 an hour, so if she teaches a seven-hour workshop, she will be paid for 14 hours.

Each classroom using the portfolio has a teacher's aide or paraprofessional for some portion of time during the day, but the cost of the aide's time is not chargeable to the portfolio, since the aides would be there anyway.

For the last year of New Jersey compensatory education funds under the present formula, Spicer arranged for a grant process so that schools could propose their own ways of using the funds, provided of course that compensatory education guidelines were followed. Each school got a minimum amount of money, but more was available to those who could propose a potentially powerful innovation. Under this program, Deans School proposed and was funded for a full-time

paraprofessional in each classroom. Constable uses its grant to pay for a full-time substitute so that teachers may be released to study and observe ways of helping at-risk children. In each case the use of the portfolio was facilitated by these compensatory education funds.

The Reading Recovery program training is expensive. The South Brunswick Public Schools has one teacher enrolled in New York University's Reading Recovery Program. She commutes for back-to-back classes in New York, and then spends the rest of her week working with children--practicing the skills she has learned in her classes--in the South Brunswick Public Schools. The teacher, formerly a high school teacher, is attached to Brunswick Acres School. The Next Century Schools grant is paying the $15,000 for the teacher to go to NYU, in addition to her salary. The grant is also paying for $7,000 in expenses (including a mandatory trip to Ohio State University) and $5,000 for the installation of a mirror window in Brunswick Acres classroom, to facilitate training in Reading Recovery.

When she returns to full-time work as a South Brunswick Public Schools teacher-leader in September 1991, she will train other teachers both in South Brunswick Public Schools and adjoining districts. These districts will pay a total of $10,000 per teacher, $9,000 for training, plus $1,000 for Rutgers University Credit. This cost includes $2500 for books and a trip to Ohio State University for a seven day Reading Recovery conference.

Section VII. Differing Impacts

Teachers believe that the portfolio is unbiased: "I don't think the portfolio looks at differences in gender, race, or SES at all. Its nondescript. It makes no difference where they're coming from, who they are. It's showing every single child's strengths and weaknesses, period...With a standardized test, you give it to everybody, there're going to some children it's not going to address. The portfolio has no preconceived notions."

The South Brunswick Public Schools portfolio measures growth over time rather than the student's ability at a single moment, so there is little possibility of "cheating" on it, and no motivation, since no rewards or sanctions are connected with it. Teachers are constantly sharing student portfolios and discrepancies are brought up to the Student Assistance Team (see discussion of Deans Elementary School for the role of the Student Assistance Team) for evaluation.

Section VIII. Discussion

Teachers experience the portfolio as a tool for individualizing instruction, not primarily as accountability. Its use has affected the physical appearance of their classrooms and the organization of their daily work, but none we observed seemed upset by the increased noise or the demands for written records of observations.

Opposition to the developmentally appropriate curriculum is dealt with by moving teachers from early childhood classes to upper grade levels if they are not comfortable with the demands of the program.

The portfolio was changed in response to teachers' criticisms. Because their concerns were addressed, there was no formal opposition at all.

Students, while unaware of the details, are conscious of an atmosphere of freedom, in which they enjoy a good deal of individual attention.

Administrators support the program in ways appropriate to their role: they provide money for paraprofessional aides, they hold information meetings for parents, they encourage and support teachers by providing opportunities for teachers to meet and attend workshops.

Parents see the portfolio used in conferences with kindergarten and grade 1 teachers. In most cases they understand it as a teacher's technique for helping their children learn, but do not expect that it will replace multiple-choice, norm-referenced testing for accountability. Parents who see the portfolio as part of instruction are pleased with its effect on their children, but one parent at least finds its "subjectivity" a threat to achievement standards.

Other personnel such as kindergarten and grade 1 aides and paraprofessionals, and high school students from the CIPED program experience the portfolio as a routine part of early childhood education.

The presence of the early childhood educational program and the portfolio assessment in the South Brunswick Public Schools are almost entirely due to the vision of Spicer. When we asked one teacher where she had first heard of these ideas, she answered "Willa." It is often the case that the success of innovations is dependent on personality and leadership, and this seems a prime example. Spicer bases innovations on research and authority, involves teachers from the beginning, and maintains open communication with all members of the educational community.

Spicer fights the battles so the teachers are able to develop the portfolio in a cocoon of support. Although she has inspired and recruited able assistants, such as Warren at Deans Elementary school who maintains an informal library of books and articles used by both teachers and parents, the structure is pivotally dependent on Spicer. Without her preliminary guardianship, the teachers would have been forced to fight the political battles and their support could have been undermined from the top.

The danger of personality-dependent programs is of course that without the leader they wilt and fade. Spicer seems to have been successful in converting the kindergarten and grade 1 teachers, and has clearly managed to transfer those in opposition, but, despite her efforts to make a self-sustaining program, there is no guarantee of success without her. However, the portfolio is becoming a fixture in the classroom that teachers rely on to help them to evaluate student progress. Spicer's interest in teacher ownership is a step in the right direction of institutionalizing the portfolio. Teachers already use it and rely on it, and the assumption can be made that as time passes it will become firmly institutionalized.

The teachers use the portfolios to individualize instruction and to organize their classrooms so that children get what they need for growth in literacy. They are grateful to have written confirmation of what they have observed, especially in conferencing with parents. One teacher summarized the use of the portfolio with parents: "If a child isn't doing as much as I'd like him to do, I can still show parents what the child is doing."

Teachers look upon assessment as information needed to modify instruction, and from that perspective they tended to be scathing about multiple-choice, norm-referenced tests. One teacher thought them "useless" from the point of view of assisting in the teaching process.

The teachers have a biassed view of assessment, understandable in their situation. They tend not to appreciate other uses of assessment for accountability, information to the public, or policy

formation.

They believe that the use of the portfolios improves their instruction and none would return to the basal-and-worksheet kind of teaching. But, as we have already noted, they have self-selected for these classroom assignments, so their comments may be biassed.

If they find the portfolio burdensome, they have channels to change it and make it workable. It is the genius of Spicer to have put the power of change into the teachers' hands, so that complaints are not whines but preludes to action. This naturally increases the teachers' responsibility, but they appeared to flourish under the burden.

However, the teachers believe that parents prefer multiple-choice test results as a yardstick of their children's progress, and that this attitude will not change soon.

"The next step is to educate the parents," said one teacher. "We need workshops for parents. That is the big issue, after we get all the teachers settled in using the portfolio. This is basically not going to be acceptable until these children get older and everyone can see that we're graduating literate kids and that's not going to be until many, many years from now."

At the same time, three of the four parents we interviewed as parents (the other parent was a board member) supported the use of the portfolio and expressed indifference at best towards the use of multiple-choice tests. It may be that the teachers are over-estimating the difficulties of changing parents' minds about assessment.

Certainly the district is sensitive to the issue. The district administration, especially Spicer, are politically wise in their approach to modifying assessment in South Brunswick Public Schools elementary schools. Spicer has added a new assessment, not replaced one with another, at the kindergarten and grade 1 level. This juxtaposition allows comparisons to be made publicly, without imposing one approach on everyone.

Given standardization and parent education (not to mention education of newspaper reporters, legislative staffs, and the business community), it is possible that a portfolio assessment could function as source of information in the slot now occupied by nationally normed multiple-choice tests. South Brunswick has amply proved that the portfolio provides feedback to instruction; for some parents it provides the information they need about their children. It does not yet provide the accountability information needed by school board members and the district administration, or the state and national departments of education.

But it could, if both teachers and public authorities were willing to concede to each other's needs: the teachers might pay a little more attention, and the public authorities a little less, to reporting in numerical terms. The essential information on a child's growth in literacy is recorded in the portfolio. Making it into useful accountability data is a question of packaging.

The exportability of the South Brunswick Public Schools portfolio assessment is potentially high. Its fame has already spread. Spicer estimates that South Brunswick Public Schools receive between 30 and 40 visitors per month to observe the portfolio program. She believes that observers come to find ideas and then incorporate what will work for them. As far as Spicer knows, no district or school has transplanted the entire portfolio program to their use.

To date, the New Jersey state education department has not responded to the South Brunswick

Public Schools observation portfolio. The fiscal uncertainties affecting the New Jersey State Department of Education have required layoffs and reorganization. The result is that this small program in a small district has not received any attention, negative or positive.

We have emphasized the involvement of teachers, the leadership of Spicer, and the sense of support the teachers experience. We should also draw attention to the innovative use of finances. While the district may in fact be no healthier financially than any other district in New Jersey, it prioritizes funds to professional development and teacher empowerment. The proposal process for the use of New Jersey compensatory education funds unleashed the creative imagination of administrators and teachers and at the same time showed that the district trusted their judgment. The district spends what one teacher called "a fortune" on new textbooks (especially the Rigby series), including some which have been developed in New Zealand. There is also an active policy of pursuing grant money to finance special programs, such as the RJR Nabisco Next Century Schools Project grant for Brunswick Acres School, and the Bruner Foundation's support of Chittenden as consultant.

Perhaps most important for the success of the early childhood instruction and assessment program is the loyalty the South Brunswick Public Schools's policies have earned from teachers. Suskin told us: "You're in unique district. It's sensational to be a teacher here--and we're not the best paid teachers, either. I wouldn't want to move for anything."

It would be impossible to say without empirical evidence that the portfolio assessment could not work in any school district, but it is reasonable to assume that in the absence of the respect for teachers that South Brunswick Public Schools demonstrates, the portfolio would not stand much chance of succeeding. This type of assessment flourishes in classrooms where teachers feel they are in control of what they do. The teachers use the portfolio, comment on it, and help to revise it. The assumption by South Brunswick Public Schools is that the teachers are the experts in the classroom, and the bottom line is that the portfolio has to work for them. In school systems where there is a different atmosphere, it would be difficult to foster the necessary cooperation between teachers and administrators. Observation-based assessment is clearly tied to developmentally appropriate instruction in early childhood; it is hard to imagine it working with basal-and-workbook instruction.

In sum, the South Brunswick Public Schools portfolio assessment serves its purposes well in its context: that is, it is an instructionally valuable tool embraced wholeheartedly by the teachers and administrators who developed it. Time and experience outside the South Brunswick Public Schools will show whether it has value beyond its present scope and has general application to other places and for other purposes.

APPENDIX A

METHODOLOGY

Arriving late Tuesday October 2, 1990, Ruth Mitchell and Amy Stempel of the Council for Basic Education and Kathleen Fulton of the Office of Technology Assessment met with Willa Spicer, Director of Curriculum and Instruction for South Brunswick Public Schools, who laid our for us our plans for the following two days. Spicer described the eight-part portfolio in detail, sketched its development over the past two years, and showed us how it is used.

During the two days of our site visit, we observed activities at three schools; met with the principals of four schools; attended Back to School Night at one school; spoke with two school board members and four parents; and extensively questioned Spicer.

Our schedule for the two-day visit:

- Morning, October 3--Deans Elementary School
- Lunch and Afternoon, October 3--Brunswick Acres Elementary School; interview at district offices with Gail Barcelo, member of the school board
- Evening, October 3--Back To School Night, Brunswick Acres Elementary School
- Morning and Lunch, October 4--Constable Elementary School
- Afternoon, October 4--Greenbrook Elementary School

CONTACTS

Willa Spicer, Director of Curriculum and Instruction
Board of Education
Four Executive Drive
Box 181, Monmouth Junction, NJ 08852
201-297-7800

John B. Wolf, Board Member
Same address
201-821-1395

Gail Barcelo, Board Member
Same address
609-921-3383

Patricia Holliday, Principal
Roberta Duffy, Joan Warren, teachers
Deans School
Georges Road, Deans
201-297-3581

Joseph Anzek, Principal
Edie Polumbo, teacher
Brunswick Acres School
Kory Drive, Kendall Park
201-297-0839

Dr. Judith Zimmerman, Principal
Sharon Suskin, Dan Diamond, teachers
Constable School
Constable Road, Kendall Park
201-297-9125

Sandy Popko, parent
Ron Woods, parent
Mr. and Ms. Philip Allsopp, parents
Stephanie Craib, principal
Greenbrook School
Roberts Street, Kendall Park
201-297-2480

APPENDIX B

Portfolio
Revised July, 1990

The committee on assessment has affirmed the purposes of maintaining a portfolio for every child who enters our program until he or she completes the second grade. First, we hope to help teachers, specialists and parents describe the child and his or her program without using standardized tests as the primary source of data. Second, the portfolio allows every teacher to document progress for every child, providing data to support and inform decisions about daily teaching and about class placement. In short, the portfolio gives substance to our contention that we take children wherever they are when they enter our program and move them forward as they become competent readers and writers.

The following list sets forth the directions for the use of the portfolio:

1. On the front of the portfolio is a "Table of Contents" covering all three years. The asterisk indicates that something should be placed in the portfolio. Many inclusions are optional but you should put a date in the square when you add an optional item.

2. Materials from each grade level should be kept in a manila folder that is clearly marked for each grade level.

3. The teacher's assessment kit includes all forms. There are many options for teachers.

   A. There are check lists available to attach to children's work samples. If you fill out a check list, staple it to the work sample and put a second asterisk on the table of contents.

   B. There are observations forms that are to be used for children who are having difficulty. If you decide to add an observation sheet, indicate so by putting an asterisk in the "other" column.

   C. There are alternative child interview forms.

   D. There are alternative forms for taking a reading sample.

4. At least once a year, each teacher will have a portfolio conference with his or her administrator or the Director of Instruction to review progress, clarify directions and talk about children who are having difficulty.

The following list sets forth the contents of the portfolio:

1. Self Portraits
   What: A picture drawn to the prompt: "Draw a picture of yourself". The picture could be drawn on the front and back of the manila folders that mark each grade level.
   When: At the beginning and end of each school year.

2. Interview with the Child
   What: The interview can take several forms depending on the level of the child. Several interview forms are included. The teacher should select the form that is most appropriate for that child.
   When: Each child should be interviewed at the beginning of each year. When you observe major changes in the behavior of the child, another interview might be in order.

3. Interview with the Parent
   What: The questionnaire found in the teacher's assessment kit should go home to parents before conference time. Discussion of the questionnaire and other related information about the child's literacy behavior should occur at the conference and notes recorded on the questionnaire form.
   When: Form goes home before fall conferences.

4. Concepts About Print Test
   What: Student test paper
   When: At the beginning and end of kindergarten. If child is having difficulty, repeat the test as needed.

5. Word Awareness Writing Activity (WAWA)
   What: Student paper marked with a score. Directions for this activity are in the assessment kit.
   When: End of Kindergarten
   Beginning and end of grade 1
   If child is having difficulty, repeat the activity as needed. (Optional Word list is provided).

6. Reading Sample
   What: Include a Reading Sample showing strategies the child is using to deal with print and a short analysis of the results. See directions for taking a reading test. Teachers may attach a descriptive check list (included in the teacher's portfolio) if more data about the student is needed.
   When: At least twice a year, until the child is using self-correcting strategies and reading fluently.

7. Writing Sample
   What: An example of free writing done by the student. It should be unedited by the teacher but the teacher can translate the child's words if they cannot be read easily. Teachers should attach a descriptive check list (found in teacher's portfolio) if more information about the child is needed.
   When: Beginning and end of every year.

8. Class Record Form
   What: A class record form which asks you to review accumulated knowledge about your students. If you answer "no" to any questions or if you do not have enough accumulated knowledge to answer at all, you should fill out an auxiliary check list about that child. If you err, it would be better to err by including too much data than not enough.
   When: A class list should be completed in fall and spring.

CHILD INTERVIEW

Name_____ Date_____

1. What is your favorite thing to do at home?

2. What's your favorite T.V. show?

3. Do you watch Sesame Street? Yes_____ No_____

4. Do you have books at home? Yes_____ No_____

5. What is your favorite book?

6. Do other people at home like to read? Yes_____ No_____

7. What do they read?

8. Does someone read to you at home?

Dear Parents or Guardians,

In South Brunswick, we hope to develop a working partnership between home and school that will be of benefit to the child. To do that, parents and teachers share their observations, concerns and expectations in many ways. One forum for the exchange of information is the teacher-parent conference held twice a year. To make the first conference most profitable, we ask you to share information about your youngster before you come to school. **Please answer the questions below and send the form back to the teacher with the child before the day of your conference.**

This practice of gathering information about the child before the first conference will occur in Kindergarten, first and second grade because in those years children change so rapidly that it is sometimes difficult to stay current. Thank you for your cooperation in this matter, I am looking forward to meeting you at our conference.

Sincerely,

Parent/Guardian's Name_____

Child's Name_____

I. My child's special interests (games, books, toys, TV programs) are:

2.What kind of reading materials are there at home?

3. Does anyone read aloud to the child?

4. Do you have any concerns about your child's language development?

5. Do you have any concerns about your child's reading development?

29

Concepts About Print Test

Administration and scoring

Before starting, thoroughly familiarise yourself with this test. Use the exact wording given below in each demonstration. (Read the instructions from the printed text for each administration.)

Say to the child, *'I'm going to read you this story but I want you to help me.'*.

Cover

Item 1 — Test: For orientation of book. Pass the booklet to the child holding the book vertically by outside edge, spine towards the child.

Say: *'Show me the front of this book.'*

Score: 1 point for the correct response.

Pages 2/3

Item 2 — Test: Concept that print, not picture, carries the message.

Say: *'I'll read this story. You help me. Show me where to start reading. Where do I begin to read?'*

Read the text.

Score: 1 for print, 0 for picture.

Pages 4/5

Item 3 — Test: For directional rules.

Say: *'Show me where to start.'*

Score: 1 for top left.

Item 4 — Say: *'Which way do I go?'*

Score: 1 for left to right.

Item 5 — Say: *'Where do I go after that?'*

Score: 1 for return sweep to left.

(Score items 3-5 if all movements are demonstrated in one response.)

Item 6 — Test: Word by word pointing.

Say: *'Point to it while I read it.'* (Read slowly, but fluently.)

Score: 1 for exact matching.

Page 6

Item 7 — Test: Concept of first and last.

Read the text.

Say: *'Show me the first part of the story.'*
'Show me the last part.'

Score: 1 point if BOTH are correct in any sense, i.e. applied to the whole text or a line, a word or a letter.

Page 7

Item 8 — Test: Inversion of picture.

Say: *'Show me the bottom of the picture'* (slowly and deliberately)

(Do NOT mention upside-down.)

Score: 1 for verbal explanation, OR, for pointing to top of page, OR, for turning the book around and pointing appropriately.

Pages 8/9

Item 9 — Test: Response to inverted print.

Say: *'Where do I begin?'*
'Which way do I go?'
'Where do I go after that?'

Score: 1 for beginning with 'The' (Sand) or 'I' (Stones), and moving right to left across the lower and then the upper line. OR 1 for turning the book around and moving left to right in the conventional manner.

Read the text now.

Pages 10/11

Item 10 — Test: Line sequence.

Say: *'What's wrong with this?'* (Read immediately the bottom line first, then the top line. Do NOT point.)

Score: 1 for comment on line order.

Pages 12/13

Item 11 — Test: A left page is read before a right page.

Say: *'Where do I start reading?'*

Score: 1 point for left page indication.

Item 12 — Test: Word sequence

Say: *'What's wrong on this page?'* (Point to the page number 12 — NOT the text.)

Read the text slowly as if it were correct.

Score: 1 point for comment on either error.

BEST COPY AVAILABLE

Item 13 Test: Letter order.

Say: '*What's wrong on this page?*' (Point to the page number 13 — NOT the text.)

Read the text slowly as if it were correct.

Score: 1 point for any ONE re-ordering of letters that is noticed and explained.

Pages 14/15

Item 14 Test: Re-ordering letters within a word.

Say: '*What's wrong with the writing on this page?*'

Read the text slowly as if it were correct.

Score: 1 point for ONE error noticed.

Item 15 Test: Meaning of a question mark.

Say: '*What's this for?*' (Point to or trace the question mark with a finger or pencil.)

Score: 1 point for explanation of function or name.

Pages 16/17

Test: Punctuation.

Read the text.

Say: '*What's this for?*'

Item 16 Point to or trace with a pencil, the fullstop (period).

Item 17 Point to or trace with a pencil, the comma.

Item 18 Point to or trace with a pencil, the quotation marks.

Item 19 Test: Capital and lower-case correspondence.

Say: '*Find a little letter like this.*'

Sand: Point to capital T and demonstrate by pointing to an upper case T and a lower case t if the child does not succeed.
Stones: As above for S and s.

Say: '*Find a little letter like this.*'

Sand: Point to capital M, H in turn.

Stones: Point to capital T, B in turn.

Score: **Sand:** 1 point if BOTH Mm and Hh are located.
Stones: 1 point if BOTH Tt and Bb are located.

Pages 18/19

Item 20 Test: Reversible words.

Read the text.

Say: '*Show me* was.'
'*Show me* no.'

Score: 1 point for BOTH correct.

Page 20

Ensure you have two pieces of light card (13 x 5cm) that the child can hold and slide easily over the line of text to block out words and letters. To start, lay the cards on the page but leave all print exposed. Open the cards out between each question asked.

Item 21 Test: Letter concepts.

Say: '*This story says* (Sand) *"The waves splashed in the hole" for* (Stones) *"The stone rolled down the hill"). I want you to push the cards across the story like this until all you can see is* (deliberately with stress) *just one letter.*' (Demonstrate the movement of the cards but do not do the exercise.)

Say: '*Now show me two letters.*'

Score: 1 point if BOTH are correct.

Item 22 Test: Word concept.

Say: '*Show me just one word.*'
'*Now show me two words.*'

Score: 1 point if BOTH are correct.

Item 23 Test: First and last letter concepts.

Say: '*Show me the first letter of a word.*'

'*Show me the last letter of a word.*'

Score: 1 point if BOTH are correct.

Item 24 Test: Capital letter concepts.

Say: '*Show me a capital letter.*'

Score: 1 point if correct.

BEST COPY AVAILABLE

Scoring Standards

Item	Pass	Score
1	Front of book.	
2	Print (not picture).	
3	Points top left at *'I took...'* (Sand); *'I walked...'* (Stones).	
4	Moves finger left to right on any line.	
5	Moves finger from the right-hand end of a higher line to the left-hand end of the next lower line, or moves down the page.	
6	Word by word matching.	
7	Both concepts must be correct, but may be demonstrated on the whole text or on a line, word or letter.	
8	Verbal explanation, or pointing to top of page, or turning the book around and pointing appropriately.	
9	Score for beginning with 'The' and moving right to left across the lower line and then the upper line, OR, turning the book around and moving left to right in the conventional movement pattern.	
10	Any explanation which implies that line order is altered.	
11	Says or shows that a left page precedes a right page.	
12	Notices at least one change of word order.	
13	Notices at least one change in letter order.	
14	Notices at least one change in letter order.	
15	Says 'Question mark', or 'A question', or 'Asks something'.	
16	Says 'Full stop', 'Period', or 'It tells you when you've said enough' or 'It's the end'.	
17	Says 'A little stop', or 'A rest', or 'A comma'.	
18	Says 'That's someone talking', 'Talking', 'Speech marks', 'Print' (from computers).	
19	Locates two capital and lower case pairs.	
20	Points correctly to both *was* and *no*.	
21	Locates one letter and two letters on request.	
22	Locates one word and two words on request.	
23	Locates both a first and a last letter.	
24	Locates one capital letter.	

Age Expectations For Items

(Age at which 50 percent of average European children pass an item. Clay, 1970)

Age / Item	5:0	5:6	6:0	6:6	7:0		5:0	5:6	6:0	6:6	7:0
1		x				13				x	
2	x					14					x
3		x				15				x	
4		x				16					x
5		x				17					x
6		x				18					x
7		x				19			x		
8		x				20			x		
9		x				21		x			
10		x				22			x		
11		x				23				x	
12				x		24				x	

DIRECTIONS
FOR
THE WORD AWARENESS WRITING ACTIVITY

This activity permits teachers to trace the development of literacy in young children. We know that when young children attempt to write words, their spelling errors reveal something about their understanding of how words are put together. The research suggests that there is a developmental pattern in the way children form words and that that pattern is exhibited by many children as they move through kindergarten and the primary grades. Children's "invented" spelling may therefore be a useful source of information for the assessment of literacy development.

Please follow the directions for giving the Word Awareness Writing Activity. Analyze the children's papers using the attached scoring guide and place the child's paper in the portfolio with a score on it.

DIRECTIONS FOR GIVING THE WORD AWARENESS WRITING ACTIVITY

This activity may be done with the whole class, small groups or individually.

1. Introduce the task by saying something like, "I'd like you to try to write some words that I will say. Even if you are not sure about what letters to use, go ahead and try." Since the task could be threatening, reassure the children that you do not expect them to know the correct spelling of the words.

2. Give each child a piece of paper (lined or unlined) and ask them to put their name at the top. (To help yourself decipher kindergartners' answers, you might ask them to fold their papers and write the first word at the top, the next just underneath, etc., moving on to the other side when there's no room at the bottom.)

3. Provide some guidance by demonstrating the activity with a practice word. "Let's try a word together. Let's write the word rat. The rat ate some cheese. What letter do you hear at the beginning of rat? What other letters might be needed?" Whatever way you choose to introduce the task, it is important to support all answers. This is not a spelling test -- we are trying to understand what children are thinking and are not concerned with their getting the right answers.

4. Dictate the rest of the words, using each in the sentence given. you may repeat each word as often as necessary. If a child seems overwhelmed and upset, you may excuse them form the activity.

1.	bed	It's time to go to bed.
2.	truck	I see a dump truck.
3.	letter	The letter is in the mailbox.
4.	bumpy	The road is bumpy.
5.	dress	I bought a new dress.
6.	jail	The thief went to jail.
7.	feet	My feet hurt.
8.	shopping	Shopping at the mall is fun.
9.	monster	The monster is scary.
10.	raced	The car raced down the road.
11.	boat	I road in a sailboat.
12.	hide	Let's play hide and seek.

SCORING THE WORD AWARENESS WRITING ACTIVITY*

A scoring chart is provided below to help you analyze the spelling. Before going further, think about the features that you will look for at each developmental level:

1. Precommunicative spelling is the "babbling" stage of spelling. Children use letters for writing words but the letters are strung together randomly. The letters in Precommunicative spelling does correspond to sounds.
   Example: OPSPS = eagle; RTAT = eighty.
2. Semiphonetic spellers know that letters represent sounds. They represent sounds with letters. Spellings are often abbreviated representing initial and/or final sounds.
   Example: E = eagle; A = eighty.
3. Phonetic spellers spell words like they sound. The speller perceives and represents all of the phonemes in a word, though spellings may be unconventional.
   Example: EGL = eagle; ATE = eighty.
4. Transitional spellers think about how words appear visually; a visual memory of spelling patterns is apparent. Spellings exhibit conventions of English orthography like vowels in every syllable and vowel digraph patterns, correctly spelled inflection endings, and frequent English letter sequences.
   Example: EGUL = eagle; EIGHTEE = eighty.
5. Correct spellers develop over years of word study and writing. Correct spelling can be categorized by instruction levels; for example, correct spelling for a body of words that can be spelled, the average fourth grader would be fourth-grade level correct spelling.

Look at the child's list. Were most of the child's spellings Precommunicative, Semiphonetic, Phonetic, Transitional, or Correct? This is the child's probable developmental level. You might feel that a child truly falls between two of the categories, but try to put in just one check mark per child. If that is difficult, think about why you are having trouble and let us know of any suggestions you might have to improve the form.

WORDS	1 PRECOMMU-NICATIVE	2 SEMIPHO-NETIC	3 PHONETIC	4 TRANSI-TIONAL	5 CORRECT
bed	random	b	bd	behd	bed
truck	random	tk	trk	truhk	truck
letter	random	lt	ldr	ledder	letter
bumpy	random	bp	bmp	bumpee	bumpy
dress	random	js	jrs	dres	dress
jail	random	jl	gal	jale	jail
feet	random	ft	fet	fete	feet
shopping	random	sp	spen	shoping	shopping
monster	random	m	mnstr	monstur	monster
raced	random	r	rast	raist	raced
boat	random	b	bot	bote	boat
hide	random	hi	hid	hied	hide

* Adapted from You Can Analyze Developmental-Spelling--and here's how to do it! by Dr. J. Richard Gentry, (Early Years, May, 1985)

OPTIONAL WORD LIST

fat (practice)

rice

stick

pet

meat

jumped

dragon

table

legs

honey

seed

bottle

cake

sink

_____

_____

_____

_____

Dates			
Context and background information about the writing: • how the writing arose • how the child went about the writing • whether the child was writing alone or with others • whether the writing was discussed with anyone while the child was working on it • kind of writing (e.g. list, letter, story, poem, personal writing, information writing) complete piece of work/extract			
Child's own response to the writing.			
Teacher's response: • to the content of the writing • to the child's ability to handle this particular kind of writing • overall impression			
Development of spelling and conventions of writing.			
What this writing shows about the child's development as a writer: • how it fits into the range of the child's previous writing • *experience/support* needed to further development			

Please keep the writing with the sample sheet

BEST COPY AVAILABLE

Instructions for taking a Reading Sample

Establish a relaxed atmosphere. Ask students to read aloud as if they were reading alone. Remind them "When you're reading and come to something that you don't know, do whatever you would do if you were reading all by yourself... as if I weren't here." Students must also know before they begin reading that when they are finished they will be asked about the passage.

I. Procedure for taking a RUNNING RECORD (For Emergent Readers)

a) Selection of material- Trade books (example:Story Box)
Using a "known" book- reveals how well the child makes use of strategies that have been taught. (use with children at very early stages)
Unknown book- reveals the child's ability to integrate strategies independently.
b) Make a copy of the text, or part of the text, and make the miscues on it.

A Running Record is a way of looking more closely at what a young child is thinking and doing as s/he reads. When analyzing a running record we are looking for growth in the following areas:
a) directional rules
b) one-to-one correspondence
c) use of visual cues (pictures, letters, words)
d) reading for meaning
e) use of reading strategies to self-correct miscues

II. Procedure for taking a MISCUE ANALYSIS (For Independent Readers)

To collect a reading sample you may choose selections from basal reading textbooks or from leveled reading inventory passages. The reading material must be unknown to the reader. The use of longer texts of approximately 150 words (basal texts or trade books) will produce more accurate information.

Have the child begin reading a passage that is at least one level below his/her current reading level. Continue reading at successively higher levels as long as the child can read independently.

When analyzing a reader's miscues we must consider the nature as well as the number of miscues. Check reading for such characteristics as:
a) using preceding context to make predictions
b) making miscues that make sense in context
c) correcting miscues that do not make sense
d) achieving a balance in the use of the syntactic, semantic and graphophonic cue systems.

Please keep in mind that the purpose of a "miscue analysis" is to determine what strategies a child is using while reading. The reading inventories should be used as general guidelines and not to establish instructional reading levels.

CLASS RECORD

Class List

Does the Child:

	attend in large and small groups?		interact in large and small groups?		retell a story ?		choose to read?		write willingly?	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring

39

STUDENT'S NAME _____

CODES: N = Not at this time S = Sometimes M = Most of the time

ORAL LANGUAGE CHECKLIST

	Pre K	Beg K	Mid K	End K	Beg 1	Mid 1	End 1	Beg 2	Mid 2	End 2
DATE										
<u>Speech Skills</u> Uses Appropriate Volume										
Articulates Clearly										
<u>Settings</u> Talks to Teachers										
Talks to Peers										
Talks in Small Group Situations										
Talks in Large Group Situations										
<u>Uses of Languages</u> Asks for Help From Peers										
Asks for Help From Adults										
Asks Questions										
Initiates Conversations										
Takes Turns Speaking										
Relates Personal Experiences										
Responds Relevently to Topic										
Expresses Point of View										
Expresses Feelings										
CAN RETELL A SEQUENCE OF EVENTS FROM EXPERIENCE										

STUDENT'S NAME _____

CODES: N = Not at this time S = Sometimes M = Most of the time

READING OBSERVATION

	Pre K	Beg K	Mid K	End K	Beg 1	Mid 1	End 1	Beg 2	Mid 2	End 2
DATE										
<u>Reading Attitude</u> Chooses to read										
Sustains interest										
Selects appropriate books										
Shares reading experience with others										
<u>Shared Book Experience (Group)</u> Listens attentively										
Joins in when able										
Responds to questions, texts, and pictures										
<u>Comprehension</u> Makes connections from own experience to what is read										
Makes predictions about text										
Gives reason for liking or disliking the book										

Student Name _____

RETELLING CHECK-LIST

Check One: Guided oral retelling _____ Date: _____
Unguided oral retelling _____ _____
Guided written retelling _____ _____
Unguided written retelling _____ _____

Book on Which Retelling Based	People	Time	Place	Problem	Plot	Resolution	Sequence	Other	Other	Prompts Needed to Guide Child	Other Guides to Help Child Retell

Checklist of Emergent / Early Reading Behavior

(Six months – 6 months later)

1. **Child can read a dictated story (Child tells: teacher records).** ☐

2. **On a running record of dictated story, child can find:**

   a word. ☐

   a first letter. ☐

   a last letter. ☐

3. **Child can name some letters in dictated story,** ☐

   e.g.

4. **During reading of dictated story, child:**

   moves left to right. ☐

   points to text (knows what is to be read). ☐

   points to individual words. ☐

   self-corrects. ☐

   uses specific strategies to solve a problem: ☐

   e.g.

5. **Knowledge of letters (both upper and lower case letters known).**

6. **Reading Sunshine Books.**

   Books (up to 3 books):

   Levels:

   Running Record:

   Accuracy rate: ☐

   Self-correction rate: ☐

   Relies on memory for reading text. ☐

   Uses other cues, e.g.

   Pictures. ☐

   Own language patterns. ☐

   Visual aspects, e.g., letters. ☐

   Checks responses using a variety of cues. ☐

   Knows that text contains the message. ☐

   Points accurately while reading. ☐

   Matches spoken words with words on the page precisely. ☐

   Knows concepts associated with direction: ☐

   beginning of book. ☐

   left to right. ☐

   return sweeps. ☐

   Searches, predicts, confirms. (Uses a process to make sense of text.) ☐

   Retells story. (Comments) ☐

SUNSHINE BOOKS

*Check all items as necessary for individual children.

STUDENT'S NAME _____

CODES: N = Not at this time S = Sometimes M = Most of the time

WRITING OBSERVATION

	Pre K	Beg K	Mid K	End K	Beg 1	Mid 1	End 1	Beg 2	Mid 2	End 2
DATE										
Attitude										
Chooses to write										
Sustains concentration										
Integrates writing into activities										
Reads writing to others										
Wants to display writing										
Picture/Print Relationship										
Draws only (may dictate)										
Draws picture, then writes										
Writes, then draws										
Editing										
Can recognize some errors (spacing, punctuation, capitals, etc.)										
Can correct some errors										

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

II. ARIZONA STUDENT ASSESSMENT PROGRAM:

A CASE STUDY OF STATEWIDE PERFORMANCE ASSESSMENT

Section I. Introduction

Arizona was selected as a site for case study for two reasons: it presented an opportunity to see a state launching a new statewide evaluation system using performance assessment, and it let us look at how a state and a test publisher interact.

Seven states have now declared a policy of moving their state educational evaluation systems towards performance assessment, gradually reducing reliance on multiple-choice, norm-referenced tests. All have approached the change differently:

- California, furthest along in the process, works with teachers and curriculum experts to design and score the assessment tasks.
- Vermont is designing a portfolio process in writing and mathematics, also involving teachers in design, administration, and review;
- Connecticut has formed a consortium with 20 states, the Coalition of Essential Schools, teachers representing the Urban Districts Leadership Consortium of the AFT, and Project Re:Learning to develop extended cooperative tasks in high school mathematics and science;
- New York has announced a policy to include performance and portfolio assessments in its extensive state testing system, probably following the immense success of its grade 4 science manipulative skills test, which has become a national model;
- Maryland, under time pressure from its legislature, has contracted with a test publisher to work with Maryland teachers and curriculum experts to develop innovative tasks including three-day integrated language arts examinations, which Maryland teachers will grade collectively;

- Kentucky has appointed a prestigious panel to oversee development of performance assessments as part of its fundamental redesign of the state's educational system; and

- Arizona has contracted with a test publisher to produce an assessment to state department specifications.

The variety of approaches stands in abrupt contrast to the uniformity of norm-referenced, multiple-choice tests, and offers to other states contemplating changes both a diversity of models and encouragement to develop these models further.

In this case study we focus on the development and implementation of the ASAP and provide preliminary feedback from the teachers it affects. We attended the three-day state conference at which the statewide Arizona Student Assessment Program (ASAP) was introduced to the school districts and personnel who will administer the assessments.

The sessions and discussion with key people were guided by these questions: What are the features of a performance assessment designed by a test publisher for a state? How did the state department of education and the publishers work together? How does a performance assessment fit into the political and educational context of the state? What do the school district personnel think of the assessment and how will they introduce it to their schools? Answers to all the questions will be found in the following discussions.

Section II. Purpose of and Rationale for the Assessment

The Arizona Student Assessment Program is a program assessment intended to assess progress towards the implementation of the Essential Skills Documents. It has two purposes: to monitor progress, and to model instructional strategies and additional content. This second purpose is unique to performance assessments. They can be deliberately used to push teaching towards desired goals because their form mirrors what teachers should be doing to implement the writing process and conceptual understanding of mathematics. For example, the ASAP reading and writing assessments are correlated so that students understand their relationship, and both extend over two days, requiring students to go through a process of revising and reflecting on their learning.

The ASAP arose from legislative concern about Arizona's educational achievement. Throughout the 1980's in the United States, this concern manifested itself in increased student testing, primarily by norm-referenced, multiple-choice tests. But also during this time industry and international student comparisons were recording a decline in the basic skills of American students. After discovering that norm-referenced, multiple-choice tests do not adequately test the desired higher order thinking skills, many states have started to search for new and different methods of assessing student progress. The ASAP is the result of just such a search.

Arizona's former statewide testing program is now ten years old; the ASAP is a new form of a familiar program. Statewide annual spring testing in grades 1-12 began in 1980 in response to the general malaise about the state of education. The tests used were the Iowa Tests of Basic Skills (ITBS) and the Tests of Achievement Proficiency (TAP), both published by the Riverside Publishing Company. The California Achievement Test (CAT) was also administered at all grade levels.

In 1987, the Arizona Legislature created a Joint Committee on Excellence in Education, a bi-partisan committee drawing from the Legislature, educators, and the public. The Committee began

by stating that Arizona needed a document that expressly told teachers, parents, and students what the State thought was important in their education. The Arizona Department of Education steered them to the already existing Essential Skills Documents. These documents began appearing in 1986, and are essentially curriculum frameworks like the ones in California. (See the case-study of the relationship between the California Assessment Program and the San Diego City Schools, case-study III.) By the end of 1990, 12 Essential Skills Documentss were available in Arizona, in the expected academic subjects plus visual and dramatic arts; dance; music; and comprehensive health. Despite their name, which gives the false impression that the documents are concerned with basic skills, the Essential Skills Documents are comprehensive descriptions of what children should know at grades 3, 8, and 12. However, the Joint Committee wanted to make sure these skills and processes were successfully being acquired by Arizona's students.

An outside, independent committee facilitated by an Arizona State University professor was hired to create a matrix and chart what items on the ITBS and the TAP were covered by the Essential Skills Documents and what items in the Essential Skills Documents were covered in the ITBS and the TAP tests. The correlations were done in mathematics, reading, and writing.

The independent committee discovered that while 100% of the items on the ITBS and TAP were covered in the Essential Skills Documents, only 20%-40%, with an average of 26%, of the Essential Skills Documents was assessed by the ITBS and the TAP. In essence, Arizona was only receiving information on how well students were mastering 26% of the content of the Essential Skills Documents and the other 74% went untested. Later correlations by Riverside confirmed the accuracy of the independent committee's work.

Some psychometricians would argue that 26% is a reliable indicator of learning. They believe that one can extrapolate from the 26% that is tested and make reliable assumptions about how a student might do on the other 74%. The Department and the Joint Committee preferred not to take the psychometric view, for two reasons: they wanted to assess as much as possible of the curriculum, and they also wanted the assessments themselves to convey a message about teaching and learning, that neither should be a matter of facts and memorization, but of understanding and application.

A teacher at the hearing when the Essential Skills Documents in mathematics was being developed said: "The teachers in Arizona can't serve two masters. If they want the teachers to do a good job of teaching math they can use the Essential Skills Documents document...and throw out the ITBS tests, or teach the ITBS tests and throw out the Essential Skills Documents" (Arizona Essential Skills for Mathematics, Arizona Department of Education, 1987, page i).

Upon the recommendations of the Joint Committee, the Legislature wrote State Law 1442 which required the Department to create assessments to adequately test the Essential Skills Documents. The bill passed by a landslide.

The Arizona Education Association (the teachers' union), the Arizona School Boards Association, the Greater Phoenix Area Curriculum Council, the Arizona School Administrators Association, and the State Chamber of Commerce were among the organizations supporting the ASAP. This diverse coalition says much about the perceived condition of education in Arizona and the state's desire to alter that perception. Even Legislators who did not know much about education were amazed at the coalition behind the ASAP and chose to support it.

Section III. Arizona's Experience to Date with the Arizona Student Assessment Program (ASAP)

Because the ASAP program is in the developmental stages, Arizona has little experience with it as an assessment, and, except for the results of piloting the assessments in the spring of 1990, little information about how teachers and students will perform. Teachers at the conference agreed that the new assessments would be harder than the norm-referenced, multiple-choice tests and that at first the students would not perform as well. They also agreed that the new assessments mirrored the activities that should go on in the classroom, requiring students strive analyze and reshape information.

As soon as State Law 1442 was passed, the Legislature wanted a program in place. The Department was given four months to produce suitable assessments. Given such a short deadline, the Department decided they could not produce assessments either themselves or with the assistance of teachers. The Department did not feel confident that the evaluation and testing divisions of state universities could provide the kind of assessment they envisioned because what they wanted differed radically from the testing norm.

The Department distributed three Requests for Quote (one each for reading, writing, and mathematics) to test publishers, with a tight deadline and $10,000 each as a ceiling for costs. The RFQ described the Department's requirements and specifications, but gave no examples of what they wanted. Riverside Publishing Company proposed a program close to what the Department had in mind--a series of innovative performance assessments. The irony is that Riverside is also the publisher of the ITBS and TAP, although different groups in the company produce the two different kinds of assessments.

The Department contracted with Riverside to develop reading, writing, and mathematics performance assessments for the benchmark ASAP grades 3, 8, and 12. These assessments are outgrowths of the Arizona Essential Skills Documents and reflect the processes and skills that educators in Arizona agree will be necessary to student success in school and in life. The assessments are standardized and are scored according to strict rubrics by Arizona teachers trained as scorers.

Riverside also developed two other less difficult versions, Forms B and C, of each assessment to be used at lower grade levels as practice assessments.

The working response to Arizona State Law 1442 is the Arizona Student Assessment Program (ASAP), which replaces the Continuous Uniform Evaluation System, (CUES), a system of district-based assessments required by state law, in place since 1972. ASAP is an umbrella program--not just the performance assessments created by Riverside Publishing Company. Under the ASAP umbrella are multiple indicators of achievement--norm-referenced, multiple choice tests; the new performance assessments; non-test indicators related to achievement; and the Arizona Report Card.

Beginning with the school year 1991-92, the ITBS and TAP will be moved to the fall and only about one third to one half of the battery will be administered each year in grades 2 through 12. Moving the tests to the fall reduces the effect Lois Easton described thus: "One curriculum till March (the one the teacher feels is most effective), then tests, then the CUES (for tests and CUES teachers must focus on the demands of the norm-referenced, multiple-choice tests), and only then back to the curriculum." The Department hopes that testing in the fall will provide a measure of what children learned in the previous year and there will be less likelihood of "teaching to the test" throughout the school year.

The Department is retaining the ITBS and the TAP for two politically sound reasons: the ITBS and the TAP are nationally norm-referenced, while performance assessments are not. The ITBS and TAP will give some idea of where Arizona's education stands in comparison to other states and will satisfy the legislature's need for information which is not filtered through the Department. The Legislature wanted to maintain some outside measure of achievement. "They essentially said, we don't trust you," Easton explained.

Home-schooling students, of which there are 2,700 approximately statewide, still take the ITBS and the TAP, but do not participate in any other part of ASAP.

Before the ASAP, the only statewide reporting occurred in the state's largest newspaper, the Arizona Republic. The paper printed charts of scores, with no analysis of their meaning in the context of the schools or those factors which affect scores--such as the percentage of students participating in the free lunch program or the percentage of students who, because their parents work in the agriculture industry, move in and out of different districts.

One of the goals of ASAP is to create report cards incorporating test information (both the ITBS and TAP tests and the performance assessments) as well as non-test indicators such as mobility rate, dropout rate, free lunches, and other factors deemed important by teachers and administrators. The Department will aggregate the various indicators and publish and distribute school, district, and state report cards.

Under ASAP, the Department is responsible for the three parts of the ASAP of the Arizona report card (ITBS and TAP, performance assessments, and non-test indicators), but districts are responsible for assessing the other grades besides 3, 8, and 12. Districts may either extend the performance assessments to the other grades, by using forms A, B, or C, or they may use other measures, including a portfolio system.

Political Climate

The ASAP arose from the Arizona Legislature's desire to measure academic achievement with valid assessments. To do this effectively a coalition had to be built and as many people as possible had to be brought on board in order for State Law 1442 to pass and the ASAP to succeed. Because of the difficulty in maintaining the delicate political balance and the necessity of the balance for ASAP's success, the political climate of Arizona is an important ingredient in the development and form of the ASAP.

Arizona became a state in 1912 and retains a sturdy independence. It is the sixth largest state in the Union, but has a population of only about three and a half million. (For comparison, Los Angeles County has more than seven million people.) One-quarter of the state's land consists of reservations, and 5% of the population is Native American. Sixteen percent is Hispanic, concentrated to some extent along the state's long border with Mexico, but present in all the state's communities. The African American population in Aizona is only 2.8%, largely found in Phoenix. There are wide contrasts in socioeconomic status in the state, from affluent communities like Scottsdale to the rural poverty of the Navajo reservations and of black and Hispanic inner-city Phoenix.

Politically, Arizona is a stronghold of conservatism: this is the state of Senator Barry Goldwater and Governor Evan Mecham. In the election which occurred in November after we visited the ASAP conference, Arizona again rejected recognizing Dr. Martin Luther King's birthday as a state holiday, although narrowly. While not directly affecting education, this is a significant indication of the

powerful forces struggling to control education in Arizona.

When we visited Arizona for the ASAP conference, the Superintendent, C. Diane Bishop, State Senator Jacque Steiner and Representative Bev Hermon were all campaigning for reelection. The three women were instrumental in the passage of the law authorizing the ASAP and are essential to its health. Superintendent Bishop beat back a challenge from Dr. Robert Miller, and Representative Hermon was easily reelected, but Senator Steiner lost. It is too early to tell how this will affect the ASAP. However, considering that two of the three were re-elected and that the person who beat Senator Steiner is not opposed to the ASAP on ideological grounds, it is doubtful that the effect will be great.

Special interest conservative groups are so well organized in Arizona that they wield an influence beyond their numerical strength. Fundamentalist Christians such as the Concerned Women of America, which has 4,000 members in Arizona, and the Christian Home School movement affect school reform.

The Department deleted an entire component of their test plan because of the fundamentalist influence. They decided not to collect information on non-test indicators related to achievement from students or parents, and not to print a student profile that would aggregate achievement and achievement-related information on a student level for use in parent-teacher conferences, because fundamentalists object to questions about home life.

The Christian home-schoolers, according to Dr. Robert Miller who ran against Superintendent Bishop for state superintendent of education, have "straight values." They do not like students to be asked to write about their feelings. Fundamentalist parents regard questions about home activities such as watching television as an invasion of privacy. They object to the phrase "realize your potential," a staple of educational rhetoric, on the grounds that "potential" implies human potential, a form of secular humanism.

It might seem strange that the home schoolers, who teach their children at home to avoid the "secular contamination" and even "Satanism" of schools, should have any interest in the system, but they say they worry about the possible spread of the ASAP to home schooling children and are concerned about the moral climate of education in Arizona.

From our experience, the opposition to the assessments was not well informed. When we interviewed Dr. Robert Miller, who was defeated for superintendent by Superintendent Bishop on 6 November 1990, he began by explaining his opposition to the science and social studies assessments. His opposition was founded on the "ambiguity" of essay tests. He stated that he "[did] not trust essay tests," and believed that these assessments would not measure the acquisition of facts. He did not seem aware, until we told him, that the science and social studies assessments have not been developed yet. We found in unnerving that, although uninformed, Dr. Miller and his constituency have great influence.

On a different level of politics, some potentially important people seemed omitted from the invitation list to the conference. Considering the amount of anxiety expressed about communicating with local district communities, it seemed surprising that local teams did not include parents, perhaps the PTA president, and school board members. Of course, we may be making a naive assumption based on experience in other states, and in fact the Department may have been avoiding trouble in the form of fundamentalist opposition by inviting only professional educators. Nevertheless, education of the public is a vital necessity if performance assessment is to gain the acceptance it needs in order to

function as intended.

Description of the assessments

To provide a flavor of these assessments, we will list the topics and tasks for all three subjects at all three levels and then describe the method in detail for one assessment, the grade eight assessment 1 in reading, writing and mathematics.

Each assessment consists of the student assessment, teacher's directions, and a scoring guide. The teacher's directions begin with a reference to the Essential Skills Document for the subject, listing the essential skill assessed, the competency indicators and the processes the student should be able to follow. The rest of the teacher's directions are essentially a script to be followed. The script includes some pre-assessment activities to focus the students on the topic.

The following is a brief overview of the assessments.

Writing: the grade three and eight prompts ask for a personal experience narrative; an imaginative story; a report based on personal observation; a communication (in the case of eighth-graders, a letter to a public figure) ; and a poem. There are nine grade 12 prompts, the first four like those of the earlier grades (the letter is a job application), with the addition of a summary; a comparison/contrast essay; a persuasive piece; and an evaluation or critique.

In each grade, the students are given the topic and do whatever is necessary to produce a rough draft--write lists of items, draw pictures (grade three), read given material, research additional information. On the second day, they rewrite, edit, and proofread their work, using an editing checklist for the final product.

The writing assessments correlate with the reading assessments. Whatever genre of communication students are asked to write, they are asked to read, and vice versa. This is an attempt to inspire integration of the language arts.

Reading: the five grade three reading assessments ask students to read and respond to a personal narrative; a story; an informative report; a communication; and a poem. The pieces are lively and interesting, and do not shy from "difficult" words such as "jesses" (in a narrative about retraining an injured owl to fly; "diplodocus dinosaur," an affectionate name shared by two friends; "captivity," in an article on play in animals.

The grade eight reading assessments repeat the grade three assessments but at a higher level, and add summary and essay. The essay is a particularly moving piece about the rehabilitation of a red-tailed hawk.

The grade 12 assessments repeat the grade eight ones and add persuasive argument and a critique, in this case a sophisticated and witty review of the movie "Star Trek V." The selections pull no punches on quality: the poem is John Donne's sonnet "Death, be not proud," and the communication is Dr. Martin Luther King's "I have a dream" speech.

Mathematics: the eight grade three mathematics assessments evaluate learning of classification; organization and interpretation of data collected by the students; simple probability; reasonable conjecture; measurement; patterns; common geometry; and translation of word problems into mathematical terms. All the assessments use narrative situations in which students play active

roles to derive answers. For example, the word problem assessment involves a map of a city with a mall as the destination for children who walk so many blocks North, South, East, or West to get to it.

The nine assessments for grade eight test knowledge of and ability to apply probability (two assessments--the one described above and predicting the results of a track meet); inductive and deductive reasoning (judging a rock-climbing contest); measuring (production of maple syrup); geometry (designing a landscape); understanding of area (working as a neighborhood odd-job person); equations with one variable (comparing two taxicab companies); extended patterns (ship's navigator); and accurate mathematical terminology (accountant for a real estate company).

At grade 12, students assume the roles of a personnel manager assessing the effectiveness of a health-care plan to show knowledge of collections and sequential procedures; analyzes data about bacterial growth to show knowledge of difference between arithmetic and geometric progressions; become "math editors" for a newspaper to identify and explain misuses of statistics; play the role of homeowners remodeling their property to show ability to apply geometry; become firefighters to estimate area, volume, and weight of common objects; conduct surveys of the student body's opinions on a class election to show ability to reason deductively and generate conclusions; act as architectural designers to show ability in nonlinear relations such as parabolas and circles; and become farmers to calculate probable yields of livestock and produce.

These more detailed examples focus on one grade 8 assessment in reading, writing, and mathematics:

Sample Reading Assessment Grade 8

This description shows how class discussion and guided process are used in the context of the assessment.

The topic of the assessment is Helen Keller and disabled people. The teacher is instructed to ask the students first : "Have you ever worked with a disabled person, for example at the Special Olympics?" The class has a few minutes to talk about their experiences with the disabled. Then the teacher asks: "If you were asked to work with a disabled person and had never had any experience, you might want to find out what it would be like. What are some questions you might want answered?" Again, discussion.

The first reading passage is a shortened and simplified version of the entry under "Keller, Helen," in the 1986 edition of the Encyclopedia Britannica. They are asked to write on a lined sheet to provide five questions "that you think a book about Helen Keller's life would answer." This portion is ungraded, but must be completed. This helps students focus their thoughts and adds to the overall quality of thee final essay. Then students read a three-page extract from Keller's The Story of My Life. They write, in the test booklet, single paragraph answers to questions about the reading: one asks them to judge whether Keller's description helped them to understand disability, and the other asked them to suggest what more might be learned from a complete reading of Keller's autobiography. Although this is a reading assessment, students must pay attention to their writing: a "Review check list" at the bottom of the page requires them to check that they have answered the questions in paragraphs containing complete sentences, checked for spelling, capitalization, punctuation, and neat writing.

The interdisciplinary emphasis is a deliberate policy to move the curriculum in the direction of

integrated learning. Each Essential Skills Documents contains indications of links to other disciplines, on the grounds that "Life is seldom parceled out the way school life is" (Designing Curriculum, p. 6). Deducing comprehension of reading through writing seems natural, but the requirement for written explanations in mathematics may be controversial.

Two further exercises are comprehension questions asking the student to characterize actions from a list of words and match up events with the sequence of the narrative. The final exercise is a one-page essay to be written in the test booklet in response to the question: "Why do you think Mr.Gibson gave the title The Miracle Worker to his play (about Anne Sullivan and Helen Keller)?" Again there is a check list at the bottom of the page. The entire series of exercises is to be completed in 45 minutes.

Sample Writing Assessment Grade 8

The topic for the grade eight writing assessment is a personal experience narrative about "an event that made you feel happy" to be included in a time capsule which will be opened by eighth graders in one hundred years. The topic (and the scoring rubric) require that there is some sense of the modern world as well as the feeling of happiness because of the intended communication to the future.

Because the invasion of privacy is one of the issues Christian fundamentalists raise about the ASAP, teachers are asked to explain that a "personal experience narrative" or autobiography must be written using the pronoun "I" but need not report the student's own feelings or experience. These may be fictional or borrowed from some other source. However, "students who choose to write about themselves have considerable latitude about what to write. They can select what is not private and still do a good job writing" (ASAP Overview, p. 26).

The 50-minute testing period on the first day is broken into 20 minutes of pre-writing and 30 minutes of drafting. The 20-minute pre-writing period is broken into five minutes of brainstorming about events that have brought happiness, all recorded in the test booklet; and 15 minutes of answers, also recorded in the booklet, to questions about the chosen experience and how it might communicate information about the late twentieth-century world to a student in 2090. (These answers are controls on the process, with the caveat mentioned above.)

When the students get their stories back the next day, there is clear assumption in the check list that they will be refining, not completely rewriting their rough drafts. They are asked to check for: correct topic; good logic; completeness; reasons why the event was important; full description; and the usual mechanical features. They write the final draft in the test booklet, and give the teacher both the rough and final drafts. The teacher's directions include the final suggestion that the students might want to read their narratives to each other in class.

The final draft is the only portion of the assessment that is scored. Scoring in done on a 4 down to 1 scale in content and organization and 4 to 1 for mechanics.

Sample Mathematics Assessment Grade 8

The mathematics assesament assesses the students' ability to find the empirical probability of an event from a sample of observed outcomes. The skills include organizing data in visual displays; using a calculator; interpreting graphs; taking polls and surveys, as well as the arithmetical and computational skills to be expected from grade eight students.

The assessment lasts for 45 minutes, which do not include a focussing session. The topic of the assessment is an election, in which the students think of themselves as campaign managers for a hypothetical Maria Clearwater (a name carefully chosen for the demographics of Arizona!) The focus questions ask students to list things which a campaign manager might have to consider to help their candidate win, such as: the candidate's rivals, and how popular they are; the voters' priorities for legislative action; and the most likely group to vote, so that they can be targeted for special attention. Again, this helps the students enter in to the assessment and contributes to the overall consistency of the student response.

The timed assessment asks students, for example, to manipulate the results of surveys, draw a bar graph, discuss the qualities of different polls, predict the percentage of voters who will turn out, and finally, write advice to Maria Clearwater on a campaign strategy: "Base your plan on the results of the research and mathematics work you have completed in this booklet. Tell her which age groups and ethnic groups of people she should aim her campaign at and why. Identify methods for choosing the issues she should focus on...."

The check list for the final essay, written in the test booklet, asks the student to check the clarity of ideas for an effective campaign strategy; make sure they are ordered logically; and carefully check for writing mechanics.

This mathematics assessment seems remarkable in its cross-disciplinary orientation. It will be instructive to find out in summer 1991 what mathematics teachers had to say about teaching political science and writing as well as mathematics, and whether the exigencies of the assessment brought about collaborations among the teachers. Certainly the assessment is intended to integrate teaching, but one wonders about the resistance to this forcible change in traditional teaching.

Section IV. Impacts on Teachers and Instruction

The ASAP is essentially a top-down program. The assessments were developed by Riverside and the Department and distributed to the teachers. It is important to note, however, that the teachers have been supported with professional development and will have the opportunity during 1990-1991 to comment on the assessments and provide suggestions to improve them. In this way, while ASAP is mandated, teachers have input into the final program.

It is difficult to evaluate the ASAP's effect on instruction at this time as 1990-1991 is a pilot year. It can be said that the ASAP evaluates student learning in a different way than the ITBS or TAP, focusing on the application of knowledge rather than the acquisition of facts. Because students are used to being evaluated by norm-referenced, multiple-choice tests and in order for students to perform well on the ASAP performance assessments, instruction will have to be altered. Arizona teacher comments indicate that much will have to change, including teacher attitudes towards their role in the classroom. We will be better able to evaluate the impact this has on instruction after the program has been in place for a few years.

The assessment forms were handed to the districts with the Department's permission and encouragement to use them for the 1990-91 school year as they please. "You have full control over what you do with them, " said Lois Easton. "That can range from trashing them to using them for practice assessments, since you get the scoring materials with them." Easton suggested the black-line masters (in that form for easy photocopying) may be used by teachers as teaching units, practice assessments, or even regular end-of-course or end-of-semester assessments; by the district as

practice versions of the assessments which will become official in 1992; as professional development material for teachers; or for informing parents at PTA meetings.

The following description of the conference will provide a detailed account of what occurred, how the ASAP was presented, and an overview of the teacher and administrator attitudes towards this new assessment program. While many of them had heard that the ASAP was coming, they had little or no idea of what was in store.

The ASAP Conference

In choosing to communicate with the school districts through a conference in the state capital, the Department was consciously following the model of California, which held the first "Beyond the Bubble" conference in October 1988 to introduce the idea of performance assessments, and then in October 1989, held two conferences, North and South, to explain what by then had become California Assessment Program policy to district and school personnel.

The timing of the ASAP conference may have been significant, in view of the elections to be held only three weeks later. The Department wished to demonstrate publicly its compliance with the Joint Committee's mandate, and the politicians, including Superintendent Bishop, wanted to rally support for education by showing off a new and promising program. The Department was taking seriously the need to keep the politicians informed about education: as Paul Koehler said in the final plenary session: "When we go to them now to talk about our schools, what they have in front of them are newspaper articles with all those numbers and all those boxes (they report only the raw ITBS, TAP, CAT, and SAT scores with no background information)--that's all they know about us, and the images aren't that good. But in every one of your schools there are wonderful things going on. Those things are not being recorded. That's what the ASAP is for--to tell them the story."

The state's 220 districts were invited to send delegates according to the number of enrolled students. A school district with fewer than 100 students was allowed two delegates, and the sliding scale went up to 10 delegates for a district with 60,000 students. The Department indicated clearly on the registration form the kind of representation they wanted from the districts: superintendents, principals, curriculum and testing coordinators, professional development coordinators, and "lead" teachers. The final count was about 850 people, judging by the crowd in the auditorium.

The program was designed to lead the school district representatives gradually from a state of suspicion and ignorance towards acceptance and trust. The first afternoon was devoted to a three-hour seminar conducted by Larry Hutchins of the Mid-Continental Regional Laboratory (McREL), which laid the foundations for change by illustrating how schools fail children and conversely what schools need to serve all students. Hutchins was especially effective in explaining in simple terms how norm-referenced, multiple-choice tests work. Before the participants left on the first day, they met in regional groups.

The ASAP was not described until the plenary session the next morning. Questions about ASAP were answered at a second plenary session on the final morning, after a day of regional meetings and break-out sessions,

The two kinds of small-group sessions were designed to educate the participants about the ASAP (these were called "break-out sessions") and to develop a regional resource support network in regional meetings.

There were 25 break-out sessions, most of them repeated three times to maximize the chances of people getting the information they needed. In addition to the sessions on the ASAP writing, reading, and mathematics assessments and their alignment with the curriculum, there were sessions on special interest areas such as Chapter 1, bilingual education, special education, and gifted students; on norm-referenced testing with the ITBS personnel; on scoring the ASAP performance assessments; and on the collection of non-test information for the Arizona report card. Some sessions looked towards the future, to integrating the curriculum and to developing the science and social studies performance assessments which are next in line if the present assessments inn writing, reading, and mathematics are successful.

The leaders of the break-out sessions included Riverside personnel, Department staff people, some school district professional development specialists, and representatives from the Center for Educational Assessment in Columbia MO, which is responsible for the organization of scoring sessions and the training of teachers to work as scorers. The sessions on the writing, reading, and mathematics performance assessments were jointly led by a Department staff person and a Riverside representative, an important message in itself.

The 17 regional meetings were each led by a Department staff member (including Lois Easton, who wished to maintain her connection with the Tucson and Maricopa County area where she worked until moving to the Department in 1987). Regions were determined by dividing the State in pie slices with Phoenix as the center. This, for the most part, allowed a mountain district to be paired with a valley district so that the different concerns could be heard by all. Regional meetings were a combination of group problemsolving and free discussion. The group leader led the participants through exercises designed to focus their attention on what needed to be done immediately upon return to their districts. There were four occasions for regions to meet during the conference.

A great deal of large-sized block paper was consumed during the break-out and regional sessions as people recorded questions about the impact of ASAP. What about the importance of writing in all these assessments? What about class size? What about transient populations--an acute problem in the Southern and border parts of the state? Would the newspapers compare the results of the ITBS and the ASAP and use the first to judge the second?

Communication with the public was perceived as a major problem. A principal said that we need ways to make ASAP clear to the public, such as charts and graphs: "We need USA Today to do the graphics."

When the 850 teachers and district personnel left the conference on 18 October, they were weighed down with black-line masters of Form A of the ASAP performance assessments in reading, writing, and mathematics for grades 3, 8 and 12. They were taking Form A back to their districts for what is essentially a statewide field test. In summer 1991, Form A will be revised again, based on the districts' feedback, and then redistributed for the same kinds of use.

In March 1991, districts will be given Form B and Form C of the assessments for field testing. Form B is intended for the non-ASAP grade immediately preceding the testing year (i.e., 2, 7, and 11) and Form C for the year preceding that (i.e., 1, 6, and 10). Form A is comprised of the practice assessments used at the ASAP grades. Forms B and C are intended to show districts how other grades should be assessed to prepare them for the ASAP and also to encourage the kind of teaching implied by the performance assessments through the grades. However, the use of forms A, B and C is not mandatory. The districts are, as we have said, responsible for assessing non-ASAP grades, and must present a plan for doing so to the Department in Spring 1991. They are offered a great

deal of help from the Department, including the option to use a portfolio system.

Form D of the performance assessments will be used for the first statewide administration of the assessments in March 1992. The content of the assessment will be secure, in the sense that the exact topics and tasks will not be revealed, but they will not be "secret," since Form D will follow the same design as that of Forms A, B, and C (the practice forms). Neither students nor teachers will face an unfamiliar situation when Form D is administered. For example, if forms A, B, and C ask students to design a playground, form D may ask them to design an amusement park. The concept and the tasks are similar but each situation is unique.

The Form A package given to the ASAP conference participants included all the assessments at all three grades: there are five assessments in reading and writing and eight in mathematics at grade three, for a total of 18 hours of either assessment or instruction; seven assessments in reading and writing and nine in mathematics at grade eight, for a total of 23 hours; and nine in reading and writing and eight in mathematics at grade 12, for a total of 26 hours. Because the performance assessments mimic instruction, they can validly function as either. Only form D is used exclusively for assessment purposes.

The quality of these assessments as tasks seems high. But questions remain about testing issues. One that seems incompletely thought out is the matrix design of the ASAP. Although all the kinds of writing will be taught, an entire class of students will be assigned one of the kinds of writing for the formal assessment. The writing test will not represent generic writing ability: it will represent writing of one genre at one particular time. Reporting will state the genre and describe it so that the parents know that the score represents writing a poem, not writing in general. Scores will be similarly specific in reading and mathematics.

One solution might be, (if it is assumed that the student received equal instruction in all the genres and therefore his/her achievement could be expected to be about the same for all genres), to have a variety of assessments administered in each class with chance accounting for which one the student faced on the booklet. Because the Arizona two-day assessment includes class discussion that option is precluded. It is also possible to arrive at a composite score across all the five or seven kinds of writing.

Certainly the issue of reporting scores has not yet been settled. Department personnel and Riverside staff are continuing to look at different options. The teachers' feedback, provided during the summer of 1991, may produce some innovative ideas. This issue will only intensify as other states and districts adopt systemically valid assessments (Maryland adopted an integrated reading/writing assessment with a three-day testing period in Fall 1990, for example).

Teachers from some of the 53 districts which piloted the Form A assessments in Spring 1990 said that they liked the assessments "because there wasn't one right answer. If they could justify the answer, that showed critical thinking." They believed that thinking will be fostered by the assessments, but not by multiple-choice. "You'll be surprised what you get out of the writing and reading on these assessments," said a third-grade teacher. "The students have to interpret what they've read, restate it, and pull together maybe five questions and an essay on it. The scoring is going to be real interesting!"

Another third-grade teacher from inner-city Phoenix said that the tests were difficult for her children because they have to follow written instructions: "they need constant verification because they're normally dependent on the teacher's direction. They would tell me they had read the

directions, and so I'd say: 'Go ahead, then do what it says.' So they'd write for a few moments, and then ask me: 'Is this what I'm supposed to do?' Because this is not what they are used to."

Twelfth-graders seemed to be looking for an answer sheet. One observer reported that they could not believe they were being asked to write in the test booklet itself. Grade 12 students have had eight years of "bubble" tests and need re-orientation.

As we attended the sessions, it became obvious that the design was working well. On the first afternoon and throughout the second day, we found people unwilling to talk to us because they didn't know enough about ASAP and what it would mean to their schools, teachers, and students. But by the third day, they felt confident and talked freely.

We suspected (and overheard) some opposition, but were unable to coax explicit statements out of people, probably because they did not want to expose their ignorance of the program, especially to us, since we had been introduced publicly as coming from Washington DC. By the third day, people had looked at examples of the ASAP, could ask intelligent questions, and began to realize that this was a pilot program, so that they could control the outcome to the extent that they offered responses and ideas to modify it.

One teacher expressed satisfaction with ASAP: "I'm real happy. I'm happy that the state has finally raised their expectations and gotten away from thinking in minimals--that was a tragedy in the past for our kids. We lag behind the majority of states in the United States. This is the first time I've seen Arizona in the lead on anything."

So did another: "We weren't helping kids, we weren't getting results with the old tests, we needed something different. This is something different." A teacher from the Yuma area, which has between 40% and 45% of its student population in English as a Second Language (ESL) classes, said that good assessments were needed for the sake of bright Mexican students, who are frequently misplaced now in mathematics. He welcomed the forthcoming Spanish version of the ASAP so that justice would be done to students whose abilities are obscured by their ESL status.

But the teachers were not uncritical. They are, said one, "positive about this, but it's going to be difficult to put into place." The difficulties include: the need for change in teaching styles; decisions about local testing, so that state and district assessments don't clash; and a massive need for professional development for teachers. "The staff development needs are incredible," said the Yuma teacher. "We need staff development on pedagogy, on writing, on logic--everything. To do this in the time frame we have, we need big bucks."

The feature of the assessments which aroused most apprehension was the primacy of writing as a learning and assessment tool. This came up frequently in the break-out sessions, the regional meetings, and in informal conversations. "Writing is our weak area," said a teacher from Fredonia, "but everything in the test depends on it." A language arts teacher added: "I worry about my son. He hates to write anything out. When he had to journal last year, he could only do it by talking into a tape-recorder. He'd rather tell you than write."

Section V. Scoring Issues

The ASAP performance assessments have not been formally scored for the entire state yet and will not be until March of 1992. The ASAP will be scored by teachers trained by the subcontractor to

Riverside, the Center for Educational Assessment Columbia, MO. Because there are many parts to each assessment, the teachers follow a scoring guide that provides the expected answers for those items that require them and rubrics to score written answers.

Many techniques are employed to maintain consistency of grading. Marked papers already read are recirculated to see if they receive the same grade again; table leaders randomly reread papers to make sure that readers are consistent; and examples of papers are made available as "anchors." We will detail the scoring techniques used during the scoring of the pilot assessments and address possible problems and concerns.

For the reading and writing assessments the scoring guide is in two parts: first, it lists expected answers for those items which require them, but with the notation "Accept equivalent phrasing," and second, rubrics for the longer questions and the final essay. These are scored from 4 at the top to 1 at the bottom, with 0 for no attempt and n/s for illegible or unreadable papers, including those written in a language other than English. It is expected that answers receiving 3 or 4 should be sufficiently free of mechanical errors (spelling, punctuation, capitalization, usage) that understanding is not impeded--but they are not expected to be perfectly "correct." "Content and creative thought are the focus of the evaluation," says the rubric for each question. There is a comprehensive description of what should be expected of a 4, 3, 2, and 1, guided by the focus on what the student has to say.

The mathematics assessments are scored in much the same way, since there are many longer response questions asking the student to explain the processes used and to draw conclusions from the data. The same grammar, spelling, usage rules apply. For the short answer questions the scoring guide provides rubrics and for those questions requiring a numerical answer, the scoring guide provides acceptable solutions. All scoring guides contain the phrases, "Accept equivalent phrasing," and "Accept equivalent fractions."

In April and May 1992, teachers will meet at regional sites for training in analytic trait scoring and for the formal scoring of the Form D assessments.

Form A was field-tested in March 1990 by 53 districts, who went through the entire process including grading sessions involving more than 100 teachers. The pilot test papers were scored at three regional centers, in Flagstaff, Phoenix, and Tucson. Riverside hired the Center for Educational Assessment to train the scoring leaders for two days before the leaders went to the regional centers and trained the teachers who were to do the scoring. Although there was a Riverside representative at each center, there seemed little need for their intervention in the process, despite the complications of maintaining standards by scoring "anchor" papers and then recirculating them unmarked. "It was much more natural than I expected," said Fred Finch of Riverside. "The teachers took right to it and the process went much better than I ever thought it would."

Extended, overnight assessments present assessment designers with the problem of standardization and reliability: how can what happens outside the test arena be controlled? If one student goes home and discusses the assignment with parents and siblings and then comes back to write a superb essay, is it fair to compare that essay with the one by a child who never thinks about the essay after the class and watches television all evening?

One control is the fact that while only the finished product is scored, the grader has access to the rough draft. A grader could detect a complete change of topic or viewpoint from first to second draft. The two drafts are intended to have a progressive relationship to each other. However, it

must be understood that to have a complete change of mind overnight is not unusual for writers of all sorts--from professional novelists who change the direction of the book when they think more about the implications of the plot, to report writers who decide on a different approach and emphasis from one draft to the next. With this in mind, any abnormal change in quality and/or vocabulary could be flagged and investigated.

When asked what the Riverside designers think about this issue (which surfaces whenever assessments include discussion, brainstorming, or work developed over more than one day), Fred Finch, chief of the design team who attended the ASAP conference, replied that the loss of standardization is the result of the "real-life" conditions of the assessment. If you are given a report to write, you may or may not take it home and work on it there, and you may or may not discuss it with a spouse or fellow workers.

There is also difficulty maintaining standards over years. A four paper one year may be a three paper the next. Efforts to control this have been tried by the California Assessment Program. CAP rescores some papers from one year to another to provide some information across the years. To date, the Arizona State Department of Education personnel have not considered this issue with the ASAP performance assessments.

These questions exemplify the problems arising from assessments which are situationally and systemically valid. There are no satisfactory answers at the moment. The questions do, however, indicate a direction for psychometric research.

Section VI. Economic Issues

The costs of the ASAP program, like those of most performance assessments, are difficult to determine because a change in assessment produces widespread and diverse effects in multiple parts of the system. The following section details the financial information available and provides a roadmap for investigating the true expenditures for assessment in the overall education budget.

Because of the one-third reduction in ITBS and TAP testing, the Department expects to use the money currently spent on the ITBS and the TAP for all three parts of the ASAP; ITBS, performance assessments, and non-test indicators. They have used Chapter II funds for the development (as opposed to the administration) of performance assessments and do not intend to ask the state legislature for any additional money for the first year or so of the program. They will undoubtedly need funding to continue to support the program after the first couple of years but there is no plan for obtaining additional money now.

The arrangement between the Department and the Riverside Publishing Company gives Riverside the copyright to these assessments. Arizona pays royalties to Riverside for the rights to unlimited copying within the state for the use of public schools only:

Form A:	Development	$73,510
	Rights of Usage	$25,255 per year
Form B:	Development	$153,585
	Rights of Usage	$25,255 per year
Form C:	Development	$153,585
	Rights of Usage	$25,255 per year

Any economic discussion must also take into account the costs of the time spent by Lois Easton, Paul Kohler, and Diane Bishop, in making policy decisions and addressing meetings on assessments; the costs of conferences introducing the assessments and their scoring; the release time for the teacher scorers to attend training sessions; and their time for scoring.

The initial conference where we observed the introduction of the ASAP cost $19,800 out of the budget, not a large cost for 850 people. But the apparent low cost disguises the fact that the school and its facilities (including two school lunches) were donated by the local school district. The participants had to pay for their own lodging for three days in motels near the school, a cost to their own districts, not to the state as such.
However, this was only one of a series of conferences: another was held in February 1991 on scoring, and others are scheduled for later this spring to discuss Forms B and C.

There will also need to be massive professional development activities for those teachers who need to change their teaching style to accommodate the new direction of Arizona's testing program. Whether that is scheduled to occur during existing professional development time using existing professional development funds or to go above and beyond existing professional development measures is unclear. By the end of the conference most teachers agreed that professional development was a priority but the decision of how to go about it was left up to individual districts.

Because the new assessments will affect curriculum and instruction, the time teachers spend revising their curriculum and teaching it should be included in the ASAP budget. As we mentioned before, the ASAP will require a comprehensive reevaluation of current pedagogy. Currently the Department does not consider this cost in the ASAP budget.

The current budget does not include the time that the state, district, and school administrators spend on decisions about how to use the conference materials and how to assess non-ASAP grades in a way that shows mastery. Because these next few years involve implementing a comprehensive new program, these administrative costs may be high.

Section VII. Differing Impacts

Statewide assessment personnel must confront equity issues when they revise the testing system. In Arizona, the major areas for equity considerations are Chapter 1, where federal regulations now require annual reporting with nationally normed tests; and American Indians, especially those living on reservations.

Chapter 1

At the ASAP conference the break-out session on the impact of ASAP on Chapter 1 raised the question of minority performance on the assessments. The participants in the session were also concerned about the shifting of the ITBS and TAP testing to the fall and whether this would make mandated reporting on Chapter I difficult. As the first year of ASAP proceeds, however, these anxieties are receding to a certain extent.

All Chapter 1 students in Arizona schools are included in the ASAP--only special education students will not participate. The ASAP is changing Chapter 1 instruction from a remedial pullout program to a revitalized experience for the students and the teachers. The teachers who are using the ASAP materials find that they give focus to their instruction, and bring Chapter 1 in line with district

goals.

For the school year 1990-91, norm-referenced data will be collected for Chapter 1 reporting from the final administration of the complete ITBS and TAP tests. In 1991-92, the fall administration of these tests will include the reading comprehension test needed for Chapter 1 reporting.

Like a number of other states, Arizona is talking to the U.S. Department of Education about using ASAP in place of nationally normed tests for Chapter 1 reporting. The department personnel are making the case that ASAP gives teachers more information about the success of students than norm-referenced tests, where one or two point gains mean little or nothing.

American Indians

The American Indian population in Arizona presents many educational and economic problems. When Easton announced to a plenary session at the conference that there would be a Spanish version of ASAP, the news was received with applause. But there was little enthusiasm for a Navajo language version, even among the personnel from Native American school districts.

The reason is that in a district like Chinle, where about 15% of the entering Navajo children do not speak English, they are illiterate in both Navajo and English, so a Navajo version of the grade three assessment would be as difficult for them as an English version.

Poor English handicaps the Navajo students even when they do not speak Navajo. They have a limited understanding of English idioms and usage. Because there is little or no economic opportunity on the reservations, education is not considered the way to a better life as it is in most other communities. Even when educated through university, Indians are torn between remaining close to their culture or leaving the reservation to find work. Many decide to stay, become bitter at the lack of opportunity, and that bitterness and desperation pervade the educational system. Many teachers who do not have ties to the community (through marriage or blood) or a strong ideological commitment to the betterment of minorities teach for only one year or even a semester before moving on to greener pastures.

While the outlook can be depressing, many innovative and ambitious educators are at work. Many hope that the ASAP program will be an improvement on the current tests which they think do not do their children justice.

Nancy Draper, Chapter I coordinator for Chinle Schools said: "I don't believe there is ever any evidence of what's really happening on our reservation because we're using ITBS. We're not satisfied with anything that's happening in evaluation at this point. More of the same is not doing anything for our kids." Motivation seems to Draper to be the major obstacle to Navajo student achievement--"they don't see any reason for it."
Another reason for low scores is the high transfer rate among schools, including the Bureau of Indian Affairs Schools: students will arrive a few days before the test and depress the scores.
A further reason is the cultural assumptions made by test publishers: "one ITBS item showed kids a banana, some yogurt, and a blender," said Sigmund Boloz principal of Ganado Primary School, on the edge of the Navajo Reservation. "Our kids don't know what yogurt is and lots of them don't know what a blender is because their homes don't have electricity."

Ganado Primary School was selected by the National Council of Teachers of English (NCTE) in 1990 as one of its five national lead schools because of its whole language program. A member of

the Arizona K-3 Task Force which developed the early-grade portions of the Essential Skills Documents, Boloz feels encouraged by the ASAP, which he approaches by seeking what it can do for the students in Ganado. He prefers to sell it to his community and teachers as an advantage for Ganado students, rather than stressing that it is a state mandate, coming from an outside group imposing its will on the school.

For six years the Ganado district has conducted a K-12 holistic writing evaluation annually, with pre-and post-instruction writing samples scored by a group of 28 teachers, so that these teachers are already trained for group scoring. Boloz actively researches evaluation results in order to show the world an accurate picture of his community, which is "so depressed," he says, that 45% of the houses have no running water and 35% have no electricity, and 95% of the students are eligible for a free lunch. An analysis of ITBS results showed that students in Ganado schools who scored above the national average did not have both parents working. He found that poor test results in Ganado were partly attributable to the transience of better educated people who could find work outside the area, leaving behind the less able, more Navajo monolingual children.
Boloz says that for his "meaning-based" school, "assessments have always gotten in the way of education....We do have to respond to Chapter 1, we do have to respond to Title VII, to groups who feel that the only way to measure growth is with an instrument that's totally inappropriate. But it doesn't tell the story."

Section VII. Discussion

The ASAP conference was a beginning. No conclusions can be drawn about the program for at least two years until the assessment has been formally administered. Our discussion of the program must be read in that light.

What distinguishes the ASAP performance assessments designed by Riverside is their appearance as a complete package. Other performance assessments so far in the United States have arrived as single subjects in varying stages of development--the writing evaluations in California for example, which began at grade eight and were developed two years later for grade 12; or the New York State grade four manipulative skills test in science, which arrived on the scene unrelated to any program of performance assessments. But test publishers like Riverside are used to presenting a product and Arizona requested a complete package.

By all accounts, the relationship between the Department and Riverside is smooth for two groups involved in an unprecedented departure from business as usual. Fred Finch said it was not the ordinary contractor-contracted agency relationship--it was a partnership from the beginning. "They understand the realities of the state of Arizona and they've listened to us about the realities of test construction. There's so much mutual respect." Of course the two organizations have worked together for years on the ITBS and the TAP, and indeed Riverside always has provided a conference for Arizona district testing personnel to explain their tests. This year, the ITBS and TAP conference was folded into the break-out sessions.

The quality of the assessments seen as curriculum material is remarkably high. The mathematics alone should cause a revolution (perhaps a revolt) in classrooms where children have traditionally filled in worksheets and memorized algorithms. The Form A assessments are perhaps more impressive as teaching material than as assessments because of the unresolved problems they present: the fuzziness about the effects of overnight reflection on two-day writing examinations and the uncertainty about deriving individual scores from the matrix design.

The Arizona legislature, led by the chairs of the Senate and House education committees, thinks of leadership in Arizona education as a cooperative affair. One politician remarked to us that Superintendent Bishop can't do anything on her own--she has to work with the Joint Committee and the Arizona State School Board. So educational improvement is inextricably connected with politics.

However, the Department showed some political savvy in providing the study that showed the discrepancy between the ITBS and the contents of the Essential Skills Documents. It made it easier to endorse performance assessments after it was demonstrated that the published tests measured an average of only 26% of the desired learning outcomes.

One member of Lois Easton's regional group called the ASAP "a voyage into a new paradigm." By the end of the conference a number of the participants seemed to understand that the assessments are not an attempt to raise scores and make people happy. They are in fact difficult, as the Phoenix teacher who wrote the grade three examination along with her students learned. But they are intended to bring about changes in teaching, curriculum, and professional attitudes.

The level of enthusiasm was uneven, depending on individual investment in the ideas supporting the ASAP. The Department's continual assurances that districts' input would be welcomed, solicited, and valued went a long way towards dissolving resistance to a top-down mandate. People went away believing themselves to be part of the process.

Since we visited, we have heard that some school districts have enthusiastically adopted the ASAP as a guide and are extrapolating its ideas to all parts of the curriculum. Easton said that she had heard of "miracles": sleepy school districts who have never discussed curriculum before have called teachers together to strategize changes. On the other hand, a recent article in the union newsletter, published by the Arizona Education Association, (an early supporter of the ASAP), complained that a few district administrators are holding on to the ASAP materials and not sharing them with teachers.

What proportion have been successful in changing their curriculum in response to the ASAP and what proportion have absorbed it, amoeba-like, into their usual routine and gone on as before will be evident in 1992. That the Legislature and the Department are giving the districts 15 months of lead time in itself testifies to their commitment. Such a thorough approach, appreciating how much time is needed for change to permeate the system, is probably due to the interdependence of the political and educational leadership in Arizona.

APPENDIX A

Ruth Mitchell and Amy Stempel of the Council for Basic Education and Michael Feuer of OTA attended the statewide ASAP conference 16 through 18 October 1990. The conference (program in the appendix) enabled us to see how the program was presented by the State Superintendent, C. Diane Bishop, and her staff, especially Paul Koehler, Assistant Superintendent, and Lois Easton, Director of Curriculum and Assessment Planning, who organized both ASAP and the conference; to interview Arizona State Senator Jacque Steiner, Chairman of the Senate Education Committee; Supt. Bishop's rival for the superintendent position, Dr. Robert Miller; Bev Hermon, Chairman of the House Education Committee; teachers and district personnel who attended the conference.

Dr. C. Diane Bishop, Superintendent
Arizona Department of Education
1535 W. Jefferson
Phoenix, AZ 85007
602-542-4361

Dr.Paul Koehler, Assistant Superintendent
same

Lois Easton, Director
Curriculum and Assessment Planning
same
602-542-5031

Representative Bev Hermon
Chair, Education Committee
House of Representatives
Capitol,
1700 West Washington
Phoenix, AZ 85007

Ms. Jacque Steiner
c/o Arizona Department of Education
1535 W. Jefferson
Phoenix, AZ 85007

Dr. Robert Miller
602-863-0574

Sigmund Boloz, Principal
Ganado Primary School
Ganado, AZ 86505
602-755-3436

Nancy Draper

Chapter I Coordinator
Chinle School District

Debra Gomez,
Third-grade teacher
Phoenix Elementary School District

Bob Brown, teacher
Yuma Union High School District 70

Fred Finch, Vice-President
Riverside Publishing Company
8420 Bryn Mawr Ave.
Chicago IL 60631
312-693-0040

APPENDIX B

Arizona
Department of Education

C. DIANE BISHOP
Superintendent

August 30, 1990

MEMORANDUM

TO: Superintendents

FROM: C. Diane Bishop, State Superintendent of Public Instruction

SUBJECT: ASAP Conference

October 16, 17 and 18 will be landmark days for education in Arizona. On those three days, state educators will gather to launch the Arizona Student Assessment Program (ASAP). The ASAP is Arizona's effort to revamp state assessment practices in order to better serve curriculum and instruction. With this project, Arizona joins about seven other states in leading the nation towards excellence in education.

SB 1442, the Goals for Educational Excellence legislation that serves as the foundation for ASAP, was signed into law by Governor Mofford on May 16. Although this law does not take effect until July 1, 1991, this coming year is an important year for piloting all parts of the law.

This is the year that districts will be able to try out the new performance-based assessments matched to the Essential Skills. This is the year that educators will help the Department write and pilot surveys of indicators related to achievement. This is the year school, district and state report cards will be drafted and reviewed by educators around the state. This is the last year the norm-referenced standardized tests will be given in the spring in their entirety.

You and the staff you designate are invited to participate in the ASAP Conference October 16, 17 and 18. You are invited to attend this conference yourself or appoint a designee. You are encouraged to form a team composed of a lead teacher, a key principal and other district staff. Your local education association may be able to help you select the lead teacher for your team.

When you leave the conference, you will take with you, in addition to all the knowledge you need about ASAP, the following tools for this pilot year:

- black-line masters of all of the new assessments;
- black-line masters of all of the scoring guides for the new assessments;

• 1535 West Jefferson, Phoenix, Arizona 85007 •
602-542-4361 FAX 602-542-5283

- a videotape that explains the ASAP for your school board, your parents and your faculty;
- a self-assessment checklist to help you prepare for ASAP; and
- a user's guide that will provide you with the information you need to implement ASAP.

Please read the enclosed brochure for details on the conference. Complete the registration form and mail it by September 14 to:

Lois Easton, Director
Curriculum and Assessment Planning
Arizona Department of Education
1535 West Jefferson
Phoenix, AZ 85007

I am looking forward to seeing you in October and working with you on this most significant effort to improve education in Arizona.

oaa511

Enclosures

HOW much will this cost?

There is no registration fee; materials are free; and some meals are provided. Delegates will need to pay for their own travel and lodging.

WHERE can we stay?

Here is a list of nearby hotels:

* Grace Inn at Ahwatukee (Best Western); 893-3000; $65.00 sgl or dbl; contact Ms. Cook.
* Wyndham Garden Hotel, 961-4444; $62.00, sgl or dbl; contact Ms. Cook.
* Inn Suites, 897-7900; $43.00/rm or $54.00/ lge suite or $50.00/king and sofa; contact Mr. Quick.
* Quality Inn, 893-3900/800-228-5151; $40.00 sgl or dbl; contact Ms. Wright.
* Comfort Inn, 820-7500/800-228-5150; $39.21 sgl or dbl; contact Ms. Stern.

When you call to make reservations, tell the contact person that you'll be attending the Arizona Department of Education ASAP Conference to get the special rates.

"In conclusion, the Commission recognizes that testing is useful and inevitable: We must know how our institutions are doing, what our children are learning, and who will make the most of opportunities not available to all. But to direct testing along a more constructive course, we must draw on richer, more direct evidence of knowledge and skill from information sources beyond multiple choice tests."

From Gatekeeper to Gateway: Transforming Testing in America. Report of the National Commission on Testing and Public Policy, 1990

HOW can we get more information?

Call:
Lois Easton, 542-5031
Barbara Nordlund, 542-5105
Judy Swanson, 542-5032

ARIZONA DEPARTMENT OF EDUCATION
1535 West Jefferson • Phoenix, AZ 85007
C. Diane Bishop
Superintendent of Public Instruction

Printed in Phoenix, AZ by the Arizona Dept. of Education
Total Copies Printed .....300 Unit Printing Cost .....$ 87
Total Printing Cost .....$261.00 Date of Printing 8/90

ASAP

Announcing a Statewide Conference on the Arizona Student Assessment Program

October 16, 17, 18, 1990
Tempe, Arizona

Arizona Department of Education
C. Diane Bishop
Superintendent of Public Instruction

BEST COPY AVAILABLE

WHO may attend?

* Districts are invited to send teams according to size of their enrollment.

Up to 100	- 2 delegates
100 to 700	- 3 delegates
700 to 1500	- 4 delegates
1500 to 5000	- 5 delegates
Over 5000	- 10 delegates

* Teams should include the superintendent (or designee) and any of the following:

  a lead teacher
  a key principal
  a curriculum coordinator
  a testing coordinator
  a staff developer

WHEN is the conference?

From 12:00 noon on Oct. 16 until 12:00 noon on Oct. 18.

WHERE is the conference?

In Ahwatukee, south of Phoenix, off I-10, at the new Mountain Point High School.
Directions: Turn east onto Warner Road off I-10 to 44th Street. Go south to Knox and east to the parking lot.

WHAT will happen?

* C. Diane Bishop, Senator Jacque Steiner, and Representative Bev Hermon will greet the delegates.
* Larry Hutchins, Mid-Continent Regional Laboratory, will provide a national perspective on Arizona's effort.
* An overview of ASAP will be presented.
* ASAP materials will be provided.
* Break-out sessions will provide detail on all aspects of ASAP.
* Regional sessions will provide assistance in implementing ASAP.

WHAT materials will we receive?

* Blackline masters of each of the new assessments.
* All materials, including scoring guides related to the new assessments.
* A videotape overview of ASAP.
* A self-assessment checklist.
* A User's Guide to ASAP.

HOW do we register?

Complete the registration form accompanying this brochure and mail it to the address on the form.

OVERVIEW
Arizona Student Assessment Program

October 1990

Arizona Department of Education
1535 W. Jefferson
Phoenix, Arizona 85007

Goals For Educational Excellence

Three years ago Arizona launched an ambitious reform effort. The Legislature, in cooperation with the State Board of Education and the Department of Education, initiated the Goals for Educational Excellence project. A Joint Legislative Committee was created to set reform goals in the areas of K-12 achievement, graduation rate, and postsecondary success in employment or higher education. In this User's Guide we deal with the K-12 achievement component.

In setting achievement goals, the Committee validated and extended Arizona's curriculum frameworks — the Essential Skills. It affirmed the Essential Skills documents as statements of its high standards for K-12 achievement. It also asked the Department to speed up the process of developing Essential Skills and to extend the development of Essential Skills beyond the required core subject areas.

With the identification of goals underway, the next question was how to measure attainment of those goals. A standardized test such as the Iowa Test of Basic Skills, which is currently in use, provides only a partial answer. One study showed that the ITBS and other standardized tests covered an average of 26% of the Essential Skills curriculum. And what it missed were the higher-order application tasks such as writing a report or using math skills for problem solving, tasks at the heart of the Essential Skills.

Recognizing this mismatch, the Committee authorized the Department to develop a high-quality, broad-based assessment system. The Department's response to this request is the Arizona Student Assessment Program. The plan calls for three kinds of state-administered assessment:

- *Essential Skills Assessments* – performance-based, curriculum-referenced assessments
- *Standardized Tests* – subtests of the IOWA Test of Basic Skills
- *Non-Test Indicators* – a survey of achievement-related factors that set the conditions for learning

State-administered assessment will include a more comprehensive and valid sampling of student learning, but it cannot tell

the full picture. It is not the state's responsibility to provide a complete measure of achievement, nor is it possible to assess all achievement through these point-in-time assessments. Instead, state assessment should be seen as a validation of more thorough, close-to-instruction district assessment. Taken together, these two sources will yield a rich array of information about student achievement.

The division of labor between school districts and the state department can be summarized as follows:

State Responsibilities:

- Essential Skills assessments at grades 3, 8, and 12, in reading, writing and mathematics
- Standardized tests given to all students (grades 2-12) but only one-third to one-half of the battery each year
- Non-test indicator surveys

District Responsibilities:

- Assessment of all students on the Essential Skills

To make assessment information accessible to a variety of audiences, a comprehensive reporting system is needed. Information will be collected from statewide testing and from school districts then communicated to educators, policy makers, and the public in Arizona Report Cards published by the Department. These Report Cards will profile achievement for individual schools, districts, and the state as a whole.

Assessment information is to be used as a basis for goal setting. Test scores and other indicators are only useful if they provide a basis for action. The plan calls for individual schools and districts, as well as the state, to examine their initial profiles and set five-year goals for improvement. Subsequent Report Cards will profile both goals set and actual scores.

The bottom line is the student. The purpose of the Goals for Educational Excellence project is to improve learning in Arizona's schools. This systematic approach will aid educators in being accountable for and responsive to student needs.

A Comparison of Current and New Assessment and Reporting

The assessment and reporting that are included in the Arizona Student Assessment Program are summarized below, with current requirements as a point of comparison.

	Current	New
State Assessment	1. Standardized Norm-Referenced Test 2. Writing Sample	1. Curriculum-Referenced Essential Skills Assessments 2. Non-Test Indicator Surveys 3. Standardized Norm-Referenced Test
State Reporting	1. Scores on Standardized Norm-Referenced Tests 2. Scores on Writing Sample	1. School Report Cards 2. District Report Cards 3. State Report Cards 4. Goal Setting At State Level
District Assessment	1. CUES (Continuous Uniform Evaluation System)	1. Selection of Options for District Assessment • Essential Skills Assessments, • District Tests, and/or • Portfolio System
District Reporting	1. None required	1. Essential Skills Completion Report 2. Goal Setting by Schools, Districts 3. Assessment Plan

Student Assessment

The Essential Skills Assessments

The new Essential Skills assessments look quite different from your typical multiple-choice standardized test. The best way to get familiar with the new assessments is to spend time looking them over. The 1990-91 school year will be well spent in this kind of examination, review, and stock-taking about the implications of the assessments for curriculum and instruction.

In this section we summarize key features of the assessments and provide some background information.

Key Points

The tests are **performance-based**. They measure student performance or application of skills, not just recall or comprehension. For instance, computation skills are assessed within the context of an applied problem.

The tests are **curriculum-referenced**. They exactly match the Essential Skills. Charts included in the Assessment Module show which skills are covered on each assessment. All skills are covered on one or more assessments. But the tests not only match the content of the Essential Skills, they also match the context of the Essential Skills. They mirror the *"how"* of the instructional processes conveyed in the documents as well as the individual skills. For example, the Language Arts Essential Skills call for doing prereading activities to prepare students for a reading selection, and this activity is included in the assessments.

The tests are **instructionally valid**. They mirror good instructional practices. For instance, the writing assessments are spaced over two days so that students may do a first draft and then come back and revise it.

The tests are **transparent**. They are not hidden away until testing day, but will be widely disseminated around the state. People should look at them, analyze what's in them, and think about them as instructional as well as assessment tools for their own use. The tests show most explicitly what we want students to be able to do, and we need to work toward that goal. This is not a problem of "teaching to the test," because the underlying processes themselves are valid. In addition, the

specific content of the items will be changed on the actual statewide testing forms (but the processes and format will be exactly parallel to the forms disseminated beforehand).

Test Specifications

Tests have been developed in reading, writing and mathematics at grades 3, 8 and 12. Additional forms of these tests will be developed soon (see Forms below). In Phase II of this project, there is an intent to develop tests in science and social studies.

Each test takes about 60 minutes to administer and is a whole, self-contained unit. For example, one test requires students to read and respond to a factual report; the entire sequence of items centers around one situation. Each test includes activities that involve teacher presentation and student oral response, often in warm-up for the scored part of the test. The tests also require teachers to give new directions at points throughout the test. For these reasons, each test will be given to a whole class of students. For example, one class will take the test on reading an informative report; another might take the test on reading a poem.

During state-administered testing, each class will do one reading assessment, one writing assessment and one mathematics assessment (with an approximate total testing time of three hours). If there is more than one class at a grade level in a school, each class will do a different assessment.

There will be four forms of each assessment:

Form A: The form that has already been developed and field-tested and will be used for the statewide pilot in 1990-91. This form will be revised one more time in the summer of 1991, based on feedback received, then redistributed to districts in black-line master format for their use. This form is targeted for grades 3, 8, and 12, but may be used by the district at any appropriate grade.

Form B: Parallel to Form A but intended for use one to two grades prior to 3, 8, and 12.

Form C: Parallel to Form A but intended for use two or more grades prior to 3, 8, and 12.

Form D: Parallel to Form A but with different content specifics. For example, where one assessment in Form A might ask children to design a playground, the parallel assessment in Form D might ask them to design an amusement park. Form D will be used for state-administered testing in the spring of 1992.

Forms B and C, the "easier" and "easiest" forms, will be distributed for statewide pilot in March of 1991, revised on the basis of feedback, and redistributed as black-line masters by September of 1992. A district could choose to use some or all of these assessments to fulfill its obligation to assess all students on the Essential Skills. (See the Assessment Module for further information.)

Test Development

The Form A tests were developed by Riverside Publishing Company based on detailed specifications and prototypes provided by the Department of Education. This process ensured that the tests would be exactly based on Arizona's Essential Skills. For example:

In reading, tests use an intact piece of real literature, good nonfiction or real language. Ungraded prereading activities are included in each assessment. Students respond to literature in a variety of ways, including writing and art work. Questions are asked to get at thinking about reading.

In writing, students do a prewriting activity and write a rough draft on day one (neither of which will be scored). They reread, revise and edit according to stated criteria on day two. Audience and purpose are established.

In mathematics, students are required to use mathematics to solve an extended problem or complete a task or project. They use manipulatives. They express understanding of mathematical concepts through writing.

All Form A tests were field tested in Arizona schools in March. Fifty-three school districts, representing all types of students and schools, participated in this field test. (See Appendix B for a list of participating districts) Over 100 Arizona educators field tested the scoring procedures in April. The results of this field test were incorporated in a revision of Form A prior to the yearlong statewide pilot.

Scoring of Assessments

Scoring the new assessments often requires more than a simple tally of right or wrong. The open-ended parts of the test will be scored through a process called analytic trait scoring. This process requires that highly specific criteria be set and that students' work be judged against these criteria. Scorers will be Arizona educators who will come together at regional scoring sites for detailed training and scoring. This process was field-tested quite successfully in April.

Norm-Referenced, Standardized Testing

The use of standardized testing is not eliminated under ASAP, just made part of a larger and more comprehensive assessment system. All students in grades 2-12 will take part of a standardized test each year.

The main purpose of standardized testing is to get an outside benchmark of how Arizona students are doing compared to students who took the test nationally. This benchmark purpose can be met satisfactorily through the administration of fewer subtests. The reduction in the number of subtests that will be given each year means that test time overall can be reduced by half to two-thirds. Most likely, the total testing time will be two hours or less.

The plan for reducing the subtests still needs to be finalized by the State Board of Education. One plan calls for a key subtest – reading comprehension – to be administered to every student every year. This subtest, of all the subtests, provides the most useful information on an ability that affects learning across the curriculum. It will be used as a "constant' while the other subtests are rotated. In addition to the subtest of read-

APPENDIX C

Chinle is the site of innovative early childhood programs for Navajo children, including a literacy center for primary grades and their parents, and a program of developing culturally relevant materials which has been given national focus in a publication by the Far West Regional Laboratory on rural programs for disadvantaged children. Dorothy King, a consultant in Native American education in Chinle, wrote a devastating account of the inadequacies of multiple-choice testing in the context of a Navajo school, entitled "Real Kids or Unreal Tasks: the Obvious Choice," which was published in the Kentucky English Bulletin, Winter 1990, and in Basic Education, October 1990.

Dorothy F. King
Box 356
Chinle, AZ 86503

KENTUCKY ENGLISH BULLETIN
Winter, 1990

REAL KIDS OR UNREAL TASKS: THE OBVIOUS CHOICE

The test results were back.

Each year, twice a year, every child from second grade through twelfth grade was given a standardized achievement test. Since progress was measured in a pre-/post- fashion, no one much worried about the pretest, but the post-test caused lots of disruption.

The students' gains or "non-gains" could be used for making decisions about stucent promotion or retention, teacher evaluation and salary increases, justification for continued funding of supplemental programs, curriculum and lesson planning, and a host of other vital phenomena.

The administrator idly looked through the class record summaries of the post-test. The second grade summaries commanded his attention. National percentile = 10 for the whole battery, grade equivalent 1.9. He'd spent a lot of time in the second grade classroom and he liked what he saw.

Perhaps, he theorized, a few extreme scores were pulling the class means down. He searched the individual scores of children he knew so well.

None of the test information made sense. A couple of students scored at the national average or above, but the others.... If one believed the scores, the second grade was a group of low functioning individuals.

86 BEST COPY AVAILABLE

This interpretation did not square with what he'd seen the children do. Obviously something was amiss.

The administrator became intrigued with the disparate information he had, with what he knew the children could do and the test scores which indicated everyone needed remediation in everything.

He strolled into the classroom. Tracy and Janelle were huddled at the computer keyboard composing a report about Japanese exports. They were discussing at which point it would be most effective to insert a chart of data base information. The teacher assistant was listening to a group of six children who planned to present a choral reading of Shel Silverstein's "Ladies First" to the first grade. Jamie was listening to Terri read CHOCOLATE FEVER--"Just skip that word, it'll make sense later," he said when she hesitated. Several children sat with their writing folders and revised pieces or used editing checklists. Mara proudly added "use quotation marks " to the Things I do well list in the back of her writing folder. Johnson was adding the word "displacement" to his personal dictionary. "Listen to me read this poem," said Darren, as he interpreted Emily Dickinson's "I'm Nobody, Who Are You?". Harold and Vance were busy measuring sticks for a building project; they recorded notations of the length and width of the sticks.

Johnny was standing on a table with binoculars looking out the window.

Everyone seemed to be busy. They certainly didn't appear to be low functioning.

It occurred to the administrator that the teacher was not in the room. He checked under a table where kids frequently curled up to read or write or listen to tapes or have private conversations. No teacher. Harold and Vance put their sticks away and joined the group of writers.

The administrator became aware that the teacher had returned and was reporting to a couple of children on an article she had read in DOLPHIN LOG, a Jacques Cousteau Society magazine.

He walked over to the group and listened to the report and the ensuing discussion. As the children left to straighten the room and leave for the day, he looked at the teacher. "The test results are back," he said. She sighed and pursed her lips. "Perhaps I'd better look at that test," he said and left.

He dug out the examiner's manual for the second grade test. "Description"..."Preparation"..."Directions"... "Test 1: Word Attack."

The students' task was to mark the word in a group of three words that had the same beginning or ending sound as a word pronounced by the examiner. The word pronounced by the examiner was not one of the words written. That meant the examiner was deliberately pronouncing a word that was not written down while children were looking at words that weren't being read to them.

The administrator knew that when people read, they often deviated from what was written, corrected the deviation if it interfered, kept going if it didn't. That was natural. But this task certainly seemed to demand very aberrant behavior from a reader and from those who were reading along.

Next task--recognizing vowel sounds, find the word among four that has the same vowel sound as a given word. That certainly sounded like a dubious life skill. Was that how poets composed? The administrator looked at the words in the sample item. He would bet that 95% of the second graders could read all those words. In fact, he was sure he could walk down the hall to their posted compositions and pick out these very words.

He rummaged for a test booklet. There was a total of sixty eight words listed in the items in this subtest. He could

BEST COPY AVAILABLE

remember children reading at least twelve of these words during his ten minute visit to the class.

With just the bearest twinge of conscience, he told himself that the school system was switching to a newer edition of the test next year and that taking this booklet out of security and showing it to the children, who had already seen it, was a necessary part of seeking the truth.

He stepped out of his office and called to a couple of second gracers who lingered on the playground. He held the pages out to them. "Can you read these words?" he asked. They looked at him patiently. "Oh, that," they said. They looked at each other and then zipped through the sixty eight words together as though they were in a race. They missed only five or six between them. "Thanks," he said. "Oh, do you know what a vowel is?" "They're letters--a,e,i,o,u, sometimes w and y. Knowing about them can help you spell, because they fit certain places in words."

"Do you have some time?" he asked. "I really need to know what you think about some things." They shrugged.

He opened the examiner's manual. Jim pointed to the heading "Reading Comprehension"; "Reading comprehension," he said. "Those stories are short and stupid," said Marie. "So are you," said Jim. Marie pointed to the word "comprehension". "That's holding what you read. When we read KOKO'S KITTEN

we found out that she could do sign language because of prehensile, a very important thing." The administrator turned more pages.

"Oh, yeah," said Jim. "This is punctuation, what part needs a capital letter. The first one." The administrator looked at the directions--decide which part needs a capital letter. The example sentence was divided into three segments. The first segment began with a capital letter. The last segment contained the uncapitalized name of a weekday. Of course, the first segment would always need a capital letter. The child had followed the directions. Directions-0, Child-1 on this task.

The administrator flipped some more pages. Sample: find the sentence that has the words in the correct order--A boy walks. Walks a boy. Boy walks a. "There's two of 'em," said Marie. "Show me," said the administrator. She pointed to the first two. "Where have you seen a sentence like that before?" asked the administrator. "At the video place--'Comes a Horseman'. Have you seen it?" "Yes," he mused, "it's one of my favorites."

He turned the page. He was losing enthusiasm; he felt he'd reached a conclusion from his research. Sample--find the answer that changes the given sentence into a question: These pens are old.--Old pens are these?--Are these pens

old?--Pens are these old? Jim pointed to "Old pens are these?" "This is like Robert Frost," he said. "What?" asked the administrator. "Teacher read us about stopping the horse in the snow," beamed Jim.

"Thanks for your help, guys." The administrator smiled as he walked off mumbling "Whose woods these are I think I know. His house is in the village though...."

The next morning, as the administrator walked to the second grade classroom, he thought about the infinite and personal nature of knowledge and the infinite possibilities, the invention of language. Perhaps, he thought, that is because knowledge and language are creations of vibrant, illimitable human minds. The pitiful bits and pieces of contrived language on the test were woefully inadequate compared to the splendor of the children's language and learning.

He entered the second grade room. He watched while one child wrote what others said on a chart labeled "Important News Items". Several children indicated that they were going to find out more about things the items caused them to think about. Others went to tables or group areas to begin reading, discussion, math.

The teacher reluctantly left the children. She walked over to the administrator. "All right. Yes, I know. You want to see me. The tests are back."

"Well, yes. The tests are back," he said. He turned to walk away. "But I can't seem to find them any more."

References:

Patterson, F. KOKO'S KITTEN. New York, NY: Scholastic, 1985.

Smith, R. K. CHOCOLATE FEVER. New York, NY: Dell, 1978.

The children described and quoted in this article ~~and~~ are bilingual and attend school in the heart of the Navajo Nation.

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

III. SAN DIEGO CITY SCHOOLS AND THE CALIFORNIA ASSESSMENT PROGRAM: A STUDY OF THEIR RELATIONSHIP

Section I. Introduction

California was the first state to declare a policy of switching its statewide assessment program--the California Assessment Program--from multiple-choice machine scorable instruments to performance assessments. The first writing examinations were implemented in 1987 at grade 8. Grade 12 was added in 1988. In 1989 the California Assessment Program began the implementation of open-ended questions in grade 12 mathematics and is now developing performance assessments in mathematics, science, history/social science, and integrated reading/writing tests. It is supporting pilot projects for portfolios in writing and mathematics.

San Diego City Schools is the second largest school district in California. It has a student population closely resembling that of major urban centers. Its personnel have been involved in the development of performance assessments in California from the beginning. The California Assessment Program and San Diego City Schools were chosen for a case-study, not of performance assessment as such, but of the political relationship between a state department of education and a large urban school district over the issue of introducing performance assessment.

Both San Diego City Schools and the California Assessment Program were in transition during 1990, largely due to budget problems. Although this might seem to disqualify them for a study of their relationship, in fact their transitional state is typical of states and districts across the country. In December 1990, efforts to document accurately the situation in the states with regard to performance assessment were thwarted by the need to define the difference between "developing" and "planning" assessments; and to categorize states which offered performance assessments although their districts ignored them and stuck with multiple-choice tests; and states where districts were using portfolios for everything from student through administrator assessment, but the state used a published multiple-choice test for accountability.

The situation in educational assessment now can best be described as a fertile chaos. Just as the California Assessment Program and San Diego City Schools are awaiting developments in the legislature, in technical expertise, and in funding, so is every other state and district concerned with change in assessment. Like the California Assessment Program and San Diego City Schools, states and districts are going ahead with ideas, plans, pilots anyway.

But while the California Assessment Program, San Diego City Schools, and their relationship reflect some nationwide concerns, their shared vision--the dovetailing of state and district assessment towards which they aspire--is uncommon. Despite differences in perspective inevitable in looking from Sacramento and from San Diego, they have essentially the same goals and they each see a role for the other in attaining their goals.

We shall attempt to characterize the relationship between San Diego City Schools and the California Assessment Program in order to understand a state-school district relationship at a critical point in the history of educational assessment in California. The characterization may provide lessons for others in similar situations.

Section II. Purpose and rationale for performance assessments in the San Diego City Schools and the California Assessment Program

Performance assessment adds a fourth purpose to the three major purposes for large-scale educational testing: determining the futures of students (tracks, special education, entrance to postsecondary education); providing feedback to teachers on the effectiveness of their instruction; and providing accountability information to legislatures and funding authorities. The fourth purpose is modeling desired instructional practices. The California Assessment Program's assessments are designed for statewide accountability and to influence teaching in new directions. Now, with the assistance of the San Diego City Schools, it is considering how to deliver individual student scores which might be used for selection and placement.

The purposes of school district testing differ: San Diego City Schools must report individual scores to parents, and aggregate scores to compare how groups of students are responding to instruction. At the same time, by law a California district must administer the California Assessment Program tests, and the results are printed in the newspapers. These different purposes, as well as other factors, could produce tension between a powerful state education department and a large progressive nationally prominent school district over the nature, frequency, and use of assessment. Each has its own perspective and priorities, which influence how they view their interaction.

The California Assessment Program: Background and Policy

The California Assessment Program has existed in the California Department of Education since 1973, when the state instituted program evaluation. The California Assessment Program found it most efficient to evaluate educational programs with matrix sampling. Its current director, Dr. Dale Carlson, has served since the beginning of the program, steering the agency skillfully through changes in demands for information and assessment techniques.

What distinguishes the California Assessment Program from similar statewide agencies is its embrace of performance assessment as a goal towards which it is moving as rapidly as limited resources will permit. In a 1989 California Education Summit manifesto, the California Assessment Program announced its policy:

> The current approach to assessment of student achievement which relies on mutliple-choice student response must be abandoned because of its deleterious effect on the educational process. An assessment system which measures student achievement on performance-based measures is essential for driving the needed reform toward a thinking curriculum in which

students are actively engaged and successful in achieving goals in and beyond high school (Final Report,p.17).

Even before this clear statement of policy, the California Assessment Program's performance assessments were being developed as a complement to the statewide curriculum frameworks, documents which began to appear in 1983 response to Senate Bill 813, California's major school reform act. The curriculum guidelines and frameworks are based on a concept of education which centralizes developing students' ability to think, to apply concepts, and to be responsible for their own learning. Writing across the curriculum, cooperative learning, experiential learning, and problemsolving are emphasized. The frameworks have become national models, adapted by many states for their own needs. The frameworks are reinforced by the California Assessment Program assessments in writing and mathematics now, and will be reinforced by the assessments in other subjects being developed.

These frameworks are not mandated, but are intended as yardsticks for school districts to measure their own curriculum and instructional techniques. However, the frameworks and guidelines are the basis of the California Assessment Program assessments, so there is an indirect pressure on school districts to align their curriculum and instruction with them. San Diego City Schools takes the frameworks so seriously that it has proposed a series of assessments, the California Framework Focused Examinations, (discussed below) to evaluate how well they are being implemented.

The pre-1990 California Assessment Program used matrix sampling, in which all students take all parts of the test but not the same items or tasks. This means that when scores are reported on the school and district level, the quality of the program can be gauged, but there are no individual student scores.

The California Assessment Program's performance assessments are what John R. Frederiksen and Allan Collins call "systemically valid," that is, they induce "in the education system curricular and instructional changes that foster the development of the cognitive traits that the test is designed to measure" ("A Systems Theory of Educational Testing," BBN Laboratories, 1990). The performance assessments were chosen by the California Assessment Program staff specifically in order to bring about changes in curriculum and instruction in the direction of the California frameworks. They thus fulfill two purposes at once--reporting on programs for accountability purposes, and motivating improvements in teaching practice.

San Diego City Schools: background and motivation

San Diego City Schools is a unified school district with close to 120,000 students, making it the second largest in California and the sixth largest city district in the U.S.

San Diego City Schools should be carefully distinguished from San Diego County Office of Education. Unlike many states in the Eastern U.S., California does not have single county-wide school districts. Los Angeles County, for example, has 90 school districts within its borders. The San Diego County Office of Education, like the Los Angeles County Office of Education, is a resource center which provides services for small school districts, but has no jurisdiction over the independent districts in the county.

There is a further peculiarity in the case of San Diego: because of political manipulation of boundaries, San Diego City covers a large geographic area, not much smaller than that of San Diego County. San Diego City Schools is therefore a large sprawling district with an urban core, but a wide range in SES from La Jolla on the Pacific Ocean to rural areas in East San Diego. San Diego

City Schools has 39% white, 16% black, 26% Hispanic, and 18% Asian students, including 8% Filipinos and 8% Vietnamese. Seventeen percent of its students need ESL instruction. Situated as it is near the border with Mexico (a trolley connects downtown San Diego with the border at Tijuana), San Diego City Schools has a large number of immigrant students, some of them undocumented.

San Diego is one of the fastest growing cities in the U.S., both from immigration and the birthrates of the Hispanic and Indochinese populations. Pressure of enrollments on San Diego City Schools has caused 32 elementary schools and one middle school to adopt year-round multiple or single-track schedules. It is under court order to desegregate the schools and offers a voluntary integration program with 45 magnet schools and 66 other schools to which students are bused free of charge.

San Diego City Schools is a dynamic school district which loses no opportunity to enrich its educational program. It is a member of the Panasonic Foundation Partnerships, a group of nine districts nationwide working with consultants from the Panasonic Foundation to restructure schools and the delivery of education; one San Diego City Schools school, Linda Vista, has a three-year grant from the RJR Nabisco Next Century Schools program; the district's middle schools are part of the Edna McConnell Clark Foundation's Program for Disadvantaged Youth; one San Diego City Schools high school offers the International Baccalaureate, with an elementary and middle school offering preparatory programs for it; San Diego City Schools has a program to train teachers in the Paideia seminar method, partially funded by the Exxon Educational Foundation; science teachers are involved in Project 2061, the program administered by the American Association for the Advancement of Science (AAAS) to redesign school science curricula.

San Diego City Schools personnel have national reputations: Superintendent Tom Payzant is a member of the National Board for Professional Teaching Standards, a trustee of the College Board, and a member of several national educational committees, including the Council of the Great City Schools. Dr. Shirley Weber, San Diego City Schools Board president, is vice-chair of the National Alliance for School Restructuring, a branch of the National Center on Education and the Economy, and Dr. Ruben Carriedo is a member of the Alliance, as well as of other national committees. A San Diego City Schools elementary teacher, Janis Gabay, was National Teacher of the Year in 1990.

San Diego City Schools maintains an open atmosphere hospitable to ideas and experiments. The district's leadership takes seriously the need for a vision to guide everyone's efforts, and a good deal of energy is devoted to revising and refining clearly stated goals to unite the disparate factions of the San Diego City Schools in improving education for all students. In 1988 Supt. Payzant convened the Innovation and Change Leadership Group (ICLG) to focus the attention of board members, administrators, and teachers' union leaders on restructuring in the broadest sense. The ICLG serves as a clearinghouse through which information about innovative practices comes into San Diego City Schools and information about district plans can be disseminated.

The encouragement of innovation is such that Superintendent Payzant has announced that only he and his deputy superintendent Bertha Pendleton can deny permission for a new program. In another notable example, the staff of a new middle school, O'Farrell Community School, was given carte blanche to design whatever innovative program they wished to try. The school began grade 7 this fall with a completely untracked instructional program. The school has a Chief Educational Officer, who performs the functions of a principal. The staff decided to trade the positions of vice principal and head counselor for extra pay so that teachers can take on these leadership responsibilities. O'Farrell staff have discussed seeking a waiver from nationally normed multiple-choice tests, because the teachers wish to concentrate on teaching not on testing. One of them characterized the O'Farrell program as "GATE for every student."

Section III. Experience to date and plans for the future.

The California Assessment Program until 1990

The California Assessment Program assessed grades 3, 6, 8, and 12 until this year, when the California Assessment Program testing became a political football. The outgoing governor of California, George Deukmejian, cut the California Assessment Program funding out of the California Department of Education's budget, and, despite a storm of protest from across the state and the nation, refused to sign either the bill to reinstate the existing program or a separate bill to support a planning process leading to development of a consolidated statewide performance assessment system. The removal of funding from the California Assessment Program was unrelated to questions of the quality of the California Assessment Program as an assessment system, but resulted from political rivalry between Governor Deukmejian and Superintendent of Public Instruction Bill Honig. For the academic year 1990-91, the California Assessment Program is proceeding with developmental activities, including some pilot-testing in San Diego City Schools.

The pause in testing has, however, provided an opportunity, albeit forced, for the California Assessment Program staff to reexamine their policies, and for a wide debate on assessment in California.

If 1990-91 had been a normal year, (although in view of the present rethinking, there never will be a "normal" year again), the California Assessment Program would have administered a grade 12 assessment in the Fall, in reading, mathematics, and writing. The reading, editing, and most of the mathematics examination are multiple-choice. The questions asked have been designed to elicit thinking, not just recognition of the correct answer, and may be called "enhanced multiple-choice." (An example can be found in the appendix.) The mathematics examination includes open-ended questions, and the writing (not the editing test) asks for an essay in one of eight genres of written communication.

In the spring, the California Assessment Program would be administering multiple-choice reading, writing, and mathematics tests to grades 3 and 6; at grade 8, multiple-choice reading, mathematics, science, and history/social science and a writing assessment asking for a written response to eight kinds of writing. As it is now (or would be if tests were being administered in 1990-91), most the California Assessment Program tests are multiple-choice, although "enhanced."

In addition to its stated policy of moving towards performance assessment, the California Assessment Program has been distinguished throughout the 1980's by its cooperation with teachers through the California Projects--the Writing Project (the model for them all, which began in Berkeley in 1974), the California Literature Project, the California Science Project, and the California Mathematics Project especially. The California Assessment Program has also worked with teachers in their professional associations, such as the California Association of Teachers of English (CATE) and the California Mathematics Council (CMC). Like the California Department of Education in general, the California Assessment Program also enlists the help of university faculty, notably Professor Charles Cooper of UC San Diego, who headed the development of the grade 8 and grade 12 writing assessments, working with groups of teachers, many in San Diego City Schools.

To get its various constituencies on board as early as possible when the performance assessment train started moving, the California Assessment Program held several conferences and symposia: Beyond the Bubble I in 1988 invited representatives of the projects, the professional organizations

and the school districts to Sacramento to hear such speakers as Professor Lauren Resnick linking assesament and a "thinking" curriculum, and Clare Burstall describing the English system of assessment. At the end of the American Educational Research Association's 1989 annual convention in San Francisco, there was a Beyond the Bubble symposium, this time addressed to the national audience attending the AERA convention. In October 1989, there were two more Beyond the Bubble conferences, one held in Sacramento (the last day coincided with the 1989 San Francisco earthquake), and the other in Long Beach. These conferences were attended by county and district administrators and explained the theory and practice of the California Assessment Program's shift to performance assessment in terms that they needed to convey the message to their constituents. A second national symposium was held at the 1990 annual convention of the AERA in Boston.

The California Assessment Program has always developed assessments by using its subject-matter consultants to coordinate the work of committees of teachers in designing, revising, and trying out tests, and now assessment tasks. The teachers are often recommended through the California Projects (mentioned above), and some Project sites (a site is a university working with adjacent school districts) undertake pilot and field-testing as part of their program.

For assessments in place, such as the grade 8 and grade 12 writing assessments, the California Assessment Program uses teachers to develop prompts and to score the essays in regional groups. A testing service, formerly the Educational Testing Service (ETS), and now Advanced Systems in Measurement and Evaluation in Dover NH (a prominent testing service in the New England states), organizes the scoring sessions, records scores, and keeps track of the essays, but in no way controls the process. The mathematics open-ended questions introduced at grade 12 in 1987 and first reported by school in 1989, are also scored by groups of teachers.

San Diego City Schools personnel have been prominent in all these statewide activities. Dr. Carriedo and Dr. Behnke have taken part in technical seminars held by the California Assessment Program, and Superintendent Payzant is chair of a statewide superintendents' committee to design the successor to the California Assessment Program. (See below.) San Diego City Schools teachers have been well represented in scoring sessions for writing and mathematics assessments, and the schools have been involved in pilot and field testing from the beginning.

<u>Plans at the state level</u>

The California Assessment Program's fallow year has provided the time necessary for new ideas to emerge. One new approach has already been embodied in a bill proposed by a California Republican senator and signed by Governor Deukmejian in August 1990. This bill provides $1 million per year for three years for two consortia of school districts, one in the northern and one in the southern part of the state, to design performance assessments that would suit their needs and possibly also those of the state. To be eligible members of the two consortia, school districts must have already experimented with some form of performance assesment, perhaps portfolios or open-ended questions.

The two consortia will be chosen by a proposal process: the California Department of Education will issue a request for proposals, review them, and make recommendations to the State Board of Education, which will award the grants. One group of school districts in Santa Clara County in the Northern part of the state has been operating as a consortium for some time. San Diego City Schools seems a natural focus for a consortium in the Southern half, especially since they have been introducing performance assessment on their own account (see below).

New legislation for statewide assessment has been introduced and Superintendent of Public Instruction Bill Honig has convened a committe consisting of superintendents of major school

districts, business representatives, university professors and teacher union representatives to advise him on an integrated assessment plan for California's educational assessment, in order to ensure a close-knit system with district support. Heading the group is Superintendent Payzant of San Diego City Schools. His appointment puts San Diego City Schools in a key position in the state's assessment future.

The charge to the superintendents' committee is the development of a comprehensive unified system for K-12 assessment, including program and individual student assessment. An integrated system would include changing the California Assessment Program grades to 4/5 (some assessment in each grade), 8, and 11, the penultimate grades in elementary, middle, and high school, so that there is a year to improve instruction before students move to the next educational level. The system would also include training teachers to score local tests using statewide standards.
Seizing the opportunity to rethink statewide performance assessment (and provide technical backup for the superintendents' committee), the California Assessment Program has convened several sessions of a the California Assessment Program Technical Planning Group to brainstorm answers to the problem. The California Assessment Program has a tradition of involving national psychometric experts in annual technical seminars, in order to ensure that the California Assessment Program assessments are sound. The present group includes members of the Center for Research on Evaluation, Standards, and Student Testing (CRESST), the national center on assessment funded by the Office of Educational Research and Improvement (OERI), of the U.S. Department of Education; educational evaluation experts from the University of California Santa Barbara, Stanford, Chicago, and Northwestern universities; representatives of ETS, Advanced Systems, Far West Laboratories (one of the regional laboratories funded by U.S. Department of Education's Office of Educational Research and Improvement), and Bolt, Beranek, and Newman of Cambridge, MA; the directors of the California Writing, Science, Mathematics, and Literature Projects; and the testing directors of California's two largest school districts, Los Angeles Unified School District, and San Diego City Schools, represented by Carriedo and Behnke.

The group is wrestling with a new challenge: maintaining the breadth of matrix-based assessment for program evaluation (and motivating instructional change) and at the same time, and if possible, with the same instruments (perhaps extended over time), providing reliable student scores on performance tasks. A model serving these two purposes would meet both state and local needs. It requires establishing performance standards to guide the design of assessment tasks and their scoring; working out the technical means of supplying individual student scores as well as matrix sampling; defining the role of test publishers, who will supply districts with tests at non-the California Assessment Program grades; and producing an assessment system which will satisfy the requirements of Chapter 1 legislation and regulations.

<u>San Diego City Schools's present situation</u>
The present San Diego City Schools districtwide mandated test is the Abbreviated Stanford Achievement Test (ASAT, marketed by the Psychological Corporation) at grades 5 and 7, administered in the spring. Before this year, a similar commercial test used to be administered at grades 5, 7, 9, and 11, so the present program represents a reduction in the number of grades tested as well as in the time devoted to testing. Any students participating in externally funded programs, such as ECIA and Chapter 1, are also tested with various levels of the ASAT in grades where the programs are operating, usually K-6. However, beginning in 1991, kindergarten students will not be tested, as part of a policy to reduce testing at the early grades.

Individual student scores are a necessary evil--"like automobiles!"--says Carriedo. San Diego City Schools needs individual and aggregatable student data to report on individual progress; to report

progress of groups by ethnicity, and residence status; to report to the court on desgregration programs; to decide on retention or promotion; and to report on students in specially funded programs such as Chapter 1 and ECIA.
Carriedo points out that San Diego City Schools especially needs to be able to aggregate by groups--which implies individual scores--in order to estimate the progress of students bused for integration. Taking La Jolla High School as an example, Carriedo said that the scores may look great, but the people on the other side of town, where the 200 black students live who are bused every day, want to see the scores for that group: "We are getting pressure now to look at groups of students, because people say that students of color are served less well."

Like the California Assessment Program, San Diego City Schools has also been rethinking its assessment system. The San Diego City Schools Testing Department, headed by Dr. Grant Behnke, has produced the New Testing System. The plans for the New Testing System make it abundantly clear that assessment in San Diego City Schools and the California Assessment Program are symbiotically related, because the New Testing System includes the California Assessment Program assessment as part of the San Diego City Schools evaluation system.

The objectives of the rethinking were to provide a flexible assessment system for the 1990s, reduce the amount of mandated testing in San Diego City Schools, introduce as much performance assessment as possible, and complement the California Assessment Program assessments to produce a streamlined testing package for the district. As a result of feasibility studies and collaboration with test publishers, children in San Diego City Schools will undergo less testing in 1990-91.

The San Diego City Schools New Testing System provides individual scores through the ASAT. However, if the California Assessment Program were sucessful in its current attempts to provide student scores as well as program evaluations, San Diego City Schools would need to rethink portions of the New Testing System.

The only districtwide testing in 1990-91 is at grades 5 and 7, although other grades will be tested for the externally funded programs. In proposing to the Board of Education the elimination of districtwide testing at grades 9 and 11, Dr. Behnke and his staff prepared a series of answers to questions about the lack of testing at higher grades (figure 1). The answers cite the inadequacies of nationally normed commercial tests for assessing higher-order thinking skills ("Conventional NRTs are not very sensitive to what is taught in secondary classrooms"), and the availability of appropriate alternatives.

These alternatives for testing at higher grades are: 1) the Golden State Examinations (GSE), also designed by the California Assessment Program, which are voluntary end-of-course examinations loosely modeled on the New York State Regents' Examinations; 2) the California Assessment Program itself, which, as a statewide program assessment, provides comparative information within California, so that San Diego City Schools knows where it stands in comparison to a statewide norm; and 3) California Framework Focused Examinations (CFFE), which are San Diego City Schools's own complement to the California Assessment Program.

The intention is to offer these California Framework Focused Examinations at non-California Assessment Program grade levels below grade 8 in order to maintain continuous alignment with the California frameworks and guidelines. The examinations were designed collaboratively by Psychological Corporation representatives and San Diego City Schools curriculum and assessment coordinators. The design, matrix sampling with a mixture of multiple-choice and open-ended items, does not seem to be entirely satisfactory yet to Behnke and Raines. However, because the tests use

Figure 1

DISTRICTWIDE TESTING PROPOSAL
"COMMONLY ASKED QUESTIONS"

1. What is the rationale for the proposal to reduce districtwide testing from grades 5, 7, 9 and 11 to grades 5 and 7? There are three main reasons: 1) One of the goals of NTS is to reduce the amount of instructional time consumed by mandated testing; 2) Conventional national norm-referenced tests (NRTs) are not very sensitive to what is taught in secondary classrooms; and, 3) The nature of NRTs (e.g., basic skills in a multiple choice format) is contradictory to the contemporary goals for secondary instruction (e.g., higher-order, creative, and divergent thinking).

2. Without the districtwide testing program at grades 9 and 11, how will we know how well our secondary students are achieving? As stated above, presently NRTs do not answer that question very well, although some people may try to use them for that purpose. Better forms of assessment are available which do respond to the question of how well students are "achieving at the secondary grade levels." The use of end-of-course exams such as the Golden State Exams (optional) is one solution. Numerous other end-of-course exams are available from professional organizations and commercial firms. In some courses, the district has prepared exams which many teachers presently use for their instructional purposes. Another excellent source of secondary grade level assessment information is the **California Assessment Program (CAP).** CAP focuses on the key curriculum objectives which should be taught in our secondary schools (California State Curriculum Frameworks). In addition, CAP provides norm-referenced information which allows us to compare our district and schools to their statewide counterparts. The new forms of CAP which have been discussed--and piloted on a limited basis--provide excellent assessment information which will be more useful for classroom instructional efforts than prior forms of testing.

3. But what if there is no CAP program, what will happen? The odds of CAP testing being available for 1991-92 are very good. Actually, there is a good likelihood that CAP testing at grades 3, 6, and 8 may still occur this spring. There was legislation introduced in the spring and summer of 1990 to permit the continuation of CAP testing and the development of promising alternate forms of assessment. Unfortunately the bills were vetoed. Assuming renewed funding of CAP, there are new forms of assessment which will occur in the near future.

4. If CAP does not return, what form of assessment will we use to measure our important "gap-closing objective" dealing with student achievement? In this time of accountability, it is extremely unlikely that CAP or some revised form of CAP would not be supported (or required) by the State legislature. If some radical departure from the present CAP evolves, then the district will need to analyze that program--as well as other alternatives--to determine how to best assess our efforts to reduce the "achievement gaps" which presently exist between schools and various subgroups of our students.

5. Why don't we just forget CAP and the political football which it has become, and develop our own assessment program? We are attempting to develop a form of assessment which will provide useful California curriculum framework information. Our limited efforts are in the form of our California Framework Focused Exams (CFFEs). These exams are intended to compliment CAP at the non-CAP grades--not to replace it. Our 1990-91 efforts have been limited to a pilot at Grade 7 only, due to budget and resource limitations. To replace a quality assessment program like CAP is not within the resources currently available in SDUSD. Trying to purchase assessment instruments from commercial firms is also a very expensive alternative--and is of questionnable worth. In both instances, the useful normative information provided by a statewide assessment program would be lost.

GB 11/90

criteria from the frameworks and guidelines, they provided preliminary information about the need to direct instruction towards them. The CFFEs were subjected to a feasibility study in grades 2, 5, and 7 in 1990. Unfortunately because of budget and staff limitations, only the grade 7 CFFE will be subjected to additional close analysis this year.

When fully developed, the CFFEs would be administered districtwide at grade 5 and available for schools to use at their own discretion and expense in other non-the California Assessment Program grade levels below grade 8. (Of course, if the California Assessment Program goes to grades 4/5, 8, and 11, then the district's plans would change.) The California Assessment Program staff are extremely interested in the CFFE and visited the training, the administration, and the scoring of the open-ended and performance tasks.

The final hypothetical question on Behnke's question-answer sheet (figure 1) is: "Why don't we just forget the California Assessment Program and the political football it has become, and develop our own assessment program?" The response highlights the value of the California Assessment Program to San Diego City Schools: "To replace a quality assessment program like the California Assessment Program is not within the resources currently available in San Diego City Schools." The CFFE examinations are intended to complement the California Assessment Program at non-California Assessment Program grades, not to replace it. Finally, commercial tests do not provide the statewide information a district gets from the California Assessment Program, and are, in Behnke's words, "of questionable worth."

Carriedo adds that the California Assessment Program has another advantage over vendor-developed tests. Vendors are concerned with making a profit and selling tests to a wider market, and cannot pay as much close attention to equity and fairness issues as is possible for the California Assessment Program.

Cooperation on future assessments

At the same time as establishing its New Testing System, San Diego City Schools has been hosting pilots for the California Assessment Program and for its own information.

In Spring 1990, San Diego City Schools administered the New York State Science Grade 4 Manipulative Skills Test, a component of the Elementary Science Program Evaluation Test (ESPET), as part of an evaluation of the district's K-6 science program. The district administered the test in 23 elementary schools, with 40 San Diego City Schools staff trained directly by the New York State Science Bureau. Since there was no money in the district budget, the participating schools paid for the training of teachers to administer the test.

The mean total score turned out to be slightly lower than that for New York fourth graders, but the value of the test, as it was in New York, lay in its message to teachers that science instruction in the early grades must equip students to be able to measure, observe, classify, infer, and interpret data, abilities which develop from hands-on experiences.

The ESPET trial was in no way a rival activity to the California Assessment Program pilot tests. It exemplifies San Diego City Schools's willingness to experiment with peformance assessments with its own resources.

San Diego City Schools pilot-tested the California Assessment Program grade 6 multiple-choice and open-ended science, and history/social studies assessments, but on a limited basis. The statewide pilot was not ready to be administered at the same time as the regular California Assessment

Program in mid-April to early May 1990. The district had already impacted elementary schools heavily with the feasibility trials for the CFFE at grades 2, 5, and 7, as well as pre-publication versions of the ASAT and did not wish to burden schools further or extend their normal testing period to accommodate the statewide pilot: "they would have been testing about the last week of school," said Bob Raines, assistant director. So San Diego City Schools said that interested teachers could voluntarily administer the California Assessment Program grade 6 assessments, and about six or seven classrooms were involved. Many other districts were piloting the assessment at the same time, so the California Assessment Program lost no information.

Carriedo expressed some disappointment that San Diego City Schools received no formal report on the results of the grade 6 the California Assessment Program pilot: "I never get a sense of which things were successful and which things failed." But the California Assessment Program could not provide feedback because of its cut in funds. In any case the grade level for this science assessment will change.

On its own account (not for the California Assessment Program), San Diego City Schools piloted a portfolio assessment in language arts with the Psychological Corporation. The instruments were developed in conjunction with the teachers and curriculum specialists of San Diego City Schools, a fact which caused Carriedo to remark that you have to do the development yourself whether you work with a publisher or not.

Thirty-two teachers of grades 1 through 8 in four elementary and three middle schools took part in the commercial portfolio project. Using reading passages supplied or suggested by the publishers, students wrote responses to prompts three times a year. These responses were scored by the Psychological Corporation and the results returned to the teachers.

At the end of the year's pilot, the teachers and the administrators agreed that they did not need to buy a commercial language arts portfolio, but could develop one themselves. As Carriedo remarked, "Portfolio is the most homegrown of all performance assessments--it has a teacher-school flavor."

Always eager to expand programs in pursuit of improvement, San Diego City Schools submitted a proposal to the Stuart Foundation this fall to fund development of performance assessments in elementary language arts and science. The proposal, which has now been funded, will provide $41,500 from the Stuart Foundation, to be matched by $46,000 in staff and teacher released time from the San Diego City Schools, and $7,500 in consultants from the Panasonic Foundation.

A proposal to the Rockefeller Foundation makes San Diego City Schools partners with the National Center for Diversified Learning and Assessment at Harvard Graduate School of Education (led by Dr. Dennis Palmer Wolf) in a middle school portfolio project. The study, which will last five years, builds on already established portfolio and other assessments in San Diego. In a literacy portfolio, students will record the early drafts and developments of the genres required for the California Assessment Program writing assessment, together with the results of classroom reading conferences. In an investigation portfolio, students will record their steps in solving problems in mathematics and science, just as some students are doing in the California Assessment Program mathematics portfolio projects. One of the objectives of the Harvard-San Diego City Schools program will be a comparison between students' portfolios with their scores on the California Assessment Program assessments, not to see which is higher, but to ascertain how much information is gained or lost by each method of assessment.

Section IV. Impacts on Teachers

Because this study focuses on the political relationship between the state and a school district, the impact on teachers is peripheral. However, the California Assessment Program performance assessments in writing have fulfilled their purpose, to model desirable forms of instruction in California classrooms.

Since the California Assessment Program grade 8 writing assessment was introduced in 1987, there has been a significant effect on writing instruction. In an evaluation of the writing assessment, the Center for the Study of Writing found that 94 percent of the California teachers surveyed assigned a greater variety of writing tasks than they did before the assessment, and 78 percent said they assigned more writing. In one year alone, the percentage of students who wrote 11 or more papers in a six-week period jumped from 22 to 33 percent.

Research in progress on open-ended mathematics questions is expected to show the same effects on instruction.

Section V. Scoring

Although scoring is not directly pertinent to the relationship between San Diego City Schools and the California Assessment Program (it has never become an issue between the two organizations), we will briefly review the major form of scoring in the California Assessment Program for general information.

When the California direct writing assessment was designed, the accepted model was a timed writing to a meaningful prompt (specifiying audience and purpose) scored by the method known as "group grading." (See Appendix B for description). The method was propagated throughout the state by the California Writing Projects, which included training in group grading in their institutes. The San Diego Writing Project was especially active in training teachers to participate, and the San Diego City Schools had a districtwide writing assessment graded by groups of teachers before the California Assessment Program introduced the statewide assessment.

There was no question of using any other kind of grading system for the California Assessment Program writing assessments, because of the connection between the writing project and the development of the assessments, both of which are teacher-oriented. The Educational Testing Service (ETS), which at the time when the writing assessments were developed, managed the logistics of the grading process, also endorsed group grading, which has been used by the Advanced Placement Program (designed and administered by ETS) since its inception.

Other states, especially Arizona and Maryland, have benefited from California's experience and built on it, so that their writing assessments reflect the writing process naturalistically: instead of a single timed writing sample, which is essentially a single draft, Arizona and Maryland writing assessments extend over two or three days, permitting pre-writing brainstorming, rough drafts, class discussion, and final editing with checklists. This is a closer approximation to the writing process as it is now taught in schools (especially those which have embraced the "whole-language" approach to literacy acquisition). However, as can be seen at a glance, while more naturalistic, the Arizona and Maryland assessments present problems in reliability and validity because of uncontrollable conditions, whereas California's simpler direct writing assessment has attained a reasonable degree of reliability.

For numbers of students higher than 300, the correlation between the 1989 and 1990 total score is .91, according to an analysis by Advanced Systems in Measurement and Evaluation, which succeeded ETS as administrators of the group grading process for the California Assessment Program writing assessments. The analysis concludes: "...The power of matrix sampling provides for highly reliable assessment, even when the number of students in a school is of modest size. Once over 50 students per grade level are tested, the reliability of the school means reaches the .90 level, and goes up from there with larger schools."

Section VI. Costs of performance assessments

Both San Diego City Schools and the California Assessment Program are suffering severely from funding crises. In fact, the California Assessment Program does not exist at this moment since its funding was cut by the governor from the education department's budget.

While the details of the budgetary problems differ in each case, the state and the district alike suffer from what they see as legislative and/or public indifference to the costs of good educational assessment. Assessment in California is seriously underfunded in comparison to the rest of the education budget. Although the sums look enormous, they in fact represent only one-twentieth of one percent of the state's education budget. Carlson has called for the "one percent solution," pointing out that such an important function deserves one percent of the budget--twenty times more than the California Assessment Program's funding until 1990.

Behnke in San Diego City Schools suffered from similar misunderstanding of the role and costs of assessment in Spring 1990, when a budget request to develop the California Framework Focused Examinations, involving teachers, curriculum resource staff, and mentor teachers was cut by the school board. "I was devastated," he said. "The message was that assessment wasn't valued...it put CFFE development on the back burner."

Carlson is repeatedly asked what performance assessment at the state level costs compared to multiple-choice tests (which the California Assessment Program has always developed and scored itself). He points out that it is impossible to compare the costs of multiple-choice machine-scorable examinations from a test publisher with a system which uses teachers to design, administer, and score tasks based on material (reflecting the frameworks) that should be taught. The budget for assessment in fact also covers professional development for teachers.

Teachers are paid for scoring writing from the assessment budget, but through participation in the training and scoring, they receive valuable professional development. "Looked at it in another way, if teachers were properly paid for staff development, the tests would be virtually free," says Carlson.

Because comparisons are not based on a solid definition of what are the costs of assessment, the figures that are cited in debates are wild and sometimes prejudiced guesses. To compare the costs accurately, it would be necessary to analyze the budget for the California Assessment Program and attribute all the costs of performance assessment to one column and the costs of multiple-choice examinations (which include parts of the present the California Assessment Program and the Golden State Examination) to another. "All the costs" include percentages of salaries devoted to one activity or the other; the costs of office staff, space, and supplies; administrative time spent making policy; contracts with testing companies for services; staff time spent in preparing publications on assessments; the costs of bringing together groups of teachers for scoring performance assessments; and the costs of scoring machine-scorable tests. Once all this information is assembled, then it

should be possible to identify costs which could be attributed to professional development and to curriculum development, which might be paid not only by the state budgets for these activities, but also by district budgets, since districts benefit from sending their teachers to regional scoring sessions.

But this thorough examination of assessment budgets has not been undertaken. We can expect unfounded assertions about the costs of performance assessments until it is.

Costs of the California Assessment Program

The California Assessment Program estimates that the writing assessment costs approximately $1.25 million per year per grade, for between 280,000 to 300,000 students. These are the costs of printing and administering the assessments and scoring them by groups of teachers organized regionally.

Because so much of the teacher involvement in the California Assessment Program is professional development (teachers call it the best professional development they have ever had), the organizational structure of SB 1882, a 1989 bill to support professional development activities, has provided regional sites and facilitated gathering groups of teachers for scoring. But no money has come from professional development funds to pay the teachers, despite the benefits they derive from the experience.

Costs of assessment in San Diego City Schools

For materials and in-house scoring, the cost of assessment in the San Diego City Schools was about $1.75 per student for the old Comprehensive Tests of Basic Skills, replaced by the Abbreviated Stanford Achievement Test. Assessment is largely financed by external sources--only 10% of San Diego City Schools's assessment money comes from the district budget. The other 90% is funded by Chapter 1, the School Improvement Program (SIP), and court-ordered integration funds which are 75% from the state. In other words, external programs which demand assessment to evaluate progress also provide the funds to carry out it out. External and district funds have been used to pay teachers for attending training to score district-developed performance assessments.

Comparing costs of machine-scorable norm-referenced published tests with district or state developed performance assessments is extremely difficult, for the reasons mentioned above. Behnke does not include the costs of his office, its personnel, equipment, and overhead, in the costs of the district's assessment. The amounts and proportions are for materials and data processing only--except in the case of external funds used for the training of teachers. Professional development funds are also used to pay teachers for scoring assessments. The California Assessment Program doesn't cost the district anything, even when San Diego City Schools teachers are trained as scorers for the writing and open-ended mathematics assessments.

But money is tight everywhere. "We are at a point in California now," says Carriedo, "where there isn't any new money. If you want to do something new, you have to get rid of something or you scale down something." If the California Assessment Program develops a method of delivering student scores as well as reports on school and district programs, it will probably seek some financial contribution from the districts.

Section VII. Impact on minority groups

The impact of performance assessments on minority groups is a serious issue in California, where more than 50% of the students in the schools are either members of minority groups or immigrants or both. There is a difference in how San Diego City Schools and the California Assessment Program personnel perceive the impacts, probably due to the relationship each has with minority groups.

The San Diego City Schools administration is accountable to a school board and president who are sympathetic to minority complaints that school districts have not treated their children equally. The school board president is eloquent about the discomfort the minority parents feel when any change in assessments is suggested. The San Diego City Schools administration does not think it can adopt a policy of eliminating multiple-choice until the minority parents and their leadership are convinced that the benefits of performance assessments outweigh the risks.

At the state level, the California Assessment Program staff deal with organizations who express different views. Bennett reports enthusiastic support from the National Coalition of ESEA-Title I Parents, Region IX. (the California organization of Chapter 1-Title 1 parents), who are largely non-white, especially in attempts to change the Chapter 1 regulations to allow performance assessments for annual reporting.

The views of the San Diego City Schools School Board President

There have been notable debates in San Diego, going back to the days of the first attempts to eliminate ability grouping in classes. San Diego City Schools has now implemented a district equity policy and a common core curriculum, which together promote equal access to all curricular programs for all students.

However, the issue of replacing multiple-choice tests with performance assessments has caused apprehension in the black community in San Diego. They suggest that the tests are being changed just when students from racially diverse backgrounds are beginning to succeed on them. Performance assessments are perceived as being subjective from design through scoring, so that parents have no assurance that their children will not be victimized by racist teachers.

The President of the San Diego City Schools Board of Education, Dr. Shirley Weber, voiced the apprehensions of the minority community at the Panasonic Partnerships Conference in Santa Fe, June 1990: "We have a long way to go to convince the public that what we're doing is in the best interests of children. We are enjoying intellectual discussions on assessment, but the reality is the public doesn't believe us," she declared. "There's another reality, that they probably shouldn't....

"When we talk about the issue of equity,...The kind of assessments we're talking about require much more faith in individuals and the belief that people can actually apply equity in testing. Most of the time with a normed test you think of something that has some subjectivity in the development of the instrument, but then in the final result you know what the answer is. When you start talking about some of the assessment we're doing--portfolios--it's all subjective. It begins subjective and it ends subjective.

"It's difficult at that point for a parent to say 'That teacher was right and that teacher's not a racist.' The teacher didn't judge my child as--my child is Black, my child is Asian, or whatever it might be. And in a district as multi-ethnic as San Diego it becomes very difficult to market to parents who don't have a level of trust about what the school district is able to do for their

children. Also the fact of course that we have a history of failure in regards to a significant number of children. Why should we believe that you've not just changed the test in order to make us believe there's some success now where there was massive failure? So we have a really big job in terms of public relations....

"... the parents are going to say, 'Uh-uh, I want to know is this A or B? I want to know, is my child above the fiftieth percentile? Is my child at grade level? Can my child read?' Now you may not think it's important but I think it's important to get my child to read."

In a telephone interview for this case-study in December 1990, Dr. Weber said that her views had not changed. She thinks that there has not been enough open dialog or staff development about the unconscious bias that she believes underlies teachers' attitudes to minority children. "Minority parents are sceptical about changing anything," she said. "They're very suspicious about what the district does."

Assistant Superintendent Carriedo commented: "Our community will not allow us to do away with multiple-choice now. We can't eliminate it until we bring them up to speed in performance assessment."

He pointed out that Dr. Weber is included as a member of the San Diego City Schools team at conferences (such as the Panasonic Partnerships conference where she spoke), and that at least one Board of Education meeting has been devoted to assessment, with Dr. Dennis Palmer Wolf of Harvard Graduate School of Education explaining portfolio assessment to the board members.

Dr. Weber is supported in her apprehensions about the subjectivity of portfolio assessments by Susan Davis, also a San Diego City Schools Board Member and participant in the Panasonic Foundation conferences. Davis thinks there is a great need for education of the community, which is confused about changes and about the status of testing since the 1990 demise of the California Assessment Program. Teachers are also confused because nothing seems settled. Nevertheless, Davis would like to see all assessments look a lot more like direct writing assessment, because, she says "Students' thought processes are much more important than their answers."

But she agrees with Dr. Weber that teachers do not respond to minority students in the same way and with the same expectations as those they have for white students. Davis said that when scores from integrated schools are disaggregated, it becomes clear that the minority students have received different treatment from their teachers and speculates about that the problem may not be the teachers in the students' present school, but in the elementary and earlier grades where they were poorly prepared for their academic future.

Davis also echoes Dr. Weber's concern with the comparability of scores. At the moment portfolios have no standardization, so it is impossible to tell whether students are performing at a level comparable to those who will compete with them.

The school board's role in encouraging change which is controversial now but will ultimately benefit all San Diego City Schools students is primarily to educate the public, says Davis. She also wants to install programs of the Parents-as-Teachers type which will help parents prepare their children for school.

She deplores the confusion surrounding assessment at the moment, but recognizes that our former confidence in norm-referenced mutiple-choice tests was misplaced. "We thought the tests were

great," she says, "but that was kinda crazy."

The California Assessment Program's perspective

The California Assessment Program exercises great care to ensure that panels of teacher-readers include representatives of the ethnic groups in the California school population. However, the disproportion between the ethnic make-up of the teaching force as opposed to that of the student population makes this difficult to achieve. Although more than 50 percent of students in California schools are members of ethnic minorities, fewer than 11 percent of teachers are either African American, Hispanic, Asian American, or Native American. This is a national problem, getting worse rather than better, because more teachers from these groups are leaving the schools than from the ranks of White teachers.

The California Assessment Program is especially involved in working to change the Chapter 1 federal reporting regulations, which worry everyone working on performance assessments. Chapter 1 regulations (which demand a nationally aggregatable test so that student improvement can be compared across states and districts) are threatening performance assessment in several states, although California has been in the forefront of attempts to change the regulations.

There are two reasons for replacing the present nationally normed multiple-choice tests: one is the effect on teaching and the other is the block that current Chapter 1 requirements put in the way of universal application of performance assessments. Chapter 1 children by definition need motivation and encouragement to engage in the kind of learning promoted by the California frameworks and guidelines. Performance assessment tasks are a means of supporting the appropriate instruction.

The needs of Chapter 1 reporting will become a major issue for the committee headed by San Diego City Schools Superintendent Payzant, as it develops an integrated system of performance assessments for the districts and the state.

Section VIII. Discussion of the relationship

The reasons for the close cooperation which is developing between the state department and the school district are: 1) financial, so that the burdens can be shared by several budgets; 2) practical, to cut down on the amount of testing and its disruptions of the classroom; 3) educational, to ensure the reinforcement of the frameworks at every grade, not just the California Assessment Program grades; and 4) administrative, to enable the California Assessment Program to maintain standards across the state.

Carriedo characterizes the California Assessment Program as "opening the door for us. They gave us permission to go ahead with performance assessment." San Diego City Schools had direct writing assessment in the early 1980s, before the California Assessment Program began its program, as part of the San Diego Writing Project. Now Carriedo thinks San Diego City Schools is pushing the California Assessment Program, hoping that it will move ahead faster in the areas, especially history/social studies, where assessments have not yet been widely piloted.

He advocates a free market in the development of assessments, so that districts could generate assessments that would not necessarily plug into the California Assessment Program: "the California Assessment Program needs to make the district folks become aware that they have the expertise--they just need to reconfigure themselves in different ways and empower themselves and take care of their own technical needs."

Behnke would also like some the California Assessment Program collaboration (and possibly money) for assessments such as the CFFEs at non-California Assessment Program grades to provide continuity, quality control, and articulation among the grade levels. He is thinking about consortia of districts working with publishers to produce these non-California Assessment Program assessments, as San Diego has. (As mentioned earlier, Assembly Bill 40, recently passed by the California legislature encourages consortia.) Behnke would like to see the California Assessment Program personnel involved in item specification in order to establish models and even pools of tasks for assessment, so that smaller districts could assess students at non-California Assessment Program grades using the California Assessment Program-approved instruments.

"We need each other--even more clearly now than in the past," said Bennett of the relationship between the California Assessment Program and the San Diego City Schools. Superintendent Honig is seeking a close relationship between the California Assessment Program and the school districts by convening the committee of which Superintendent Payzant of San Diego City Schools is the chair. The impulses towards an integrated system are strong, but technical issues have not yet been worked out.

Both San Diego City Schools and the California Assessment Program are pushing each other towards performance assessments, but they sometimes are driven by different goals and priorities, so that their efforts don't reinforce each other. For its own reasons, San Diego City Schools found it impossible to pilot-test the California Assessment Program grade 6 science and social science assessments in Spring 1990, despite their expressed desire for development in these subjects. On the other hand, no matter how hard the state pushes towards performance assessments, San Diego City Schools has the job of convincing the local community that they are better than multiple-choice--not feasible at the moment.

At the moment, most of the assessments used by the California Assessment Program and by San Diego City Schools are still multiple-choice. Despite the enthusiasm expressed by Carriedo and Behnke, in San Diego City Schools, and Carlson and Bennett, at the California Assessment Program, it is not easy to implement performance assessments as quickly or as fully as they would like.

The problems in San Diego City Schools are rooted in finances and in community relations. Money is tight everywhere, and a school district like San Diego City Schools which wants to try new programs and projects has to make difficult choices. The community attitudes expressed by Dr. Weber make a radical move to performance assessment impossible without careful demonstration that this form of assessment will not harm students who, they believe, have never been well served by the school system.

The California Assessment Program's policy differs from San Diego's policy in the California Assessment Program's intention to eliminate multiple-choice tests as soon as development of alternatives allows the change. But the California Assessment Program is luckier than San Diego City Schools--it does not have a constituency on which it depends that supports traditional tests. Its loyalty is primarily to the state curriculum frameworks and to reinforcing their implementation.

The relationship of the California Assessment Program and San Diego City Schools in the past differs from what it will have to be in the future. The past relationship was based on two separate systems. The California Assessment Program had one set of imperatives, and the San Diego City Schools, like all California districts, had to comply but also satisfy their own local demands.

The future relationship will have to balance state requirements and local district needs in an integrated statewide system. This will mean a constant free exchange of viewpoints and a commitment to communication. However, because the San Diego City Schools Superintendent, Tom Payzant, is the chair of the statewide superintendents' committee advising Supt. Honig on assessment, and both Carriedo and Behnke are regular members of the California Assessment Program Technical Planning Group, the future relationship will probably be even closer than before. San Diego City Schools and the California Assessment Program will share the leadership in developing a system of performance assessments with essential roles for the state department, school districts, and test publishers.

APPENDIX A

VISITS AND CONTACTS,

SAN DIEGO CITY SCHOOLS AND THE California Assessment Program

Ruth Mitchell met with Linda Carstens, Dr. Ruben Carriedo, Dr. Grant Behnke, and Dr. Robert Raines at the administrative offices of San Diego City Schools for formal interviews on 28 November 1990. She interviewed Dr. Sue Bennett by telephone on 27 December 1990. These contacts, however, should be seen against the background of a long association of Mitchell with both San Diego City Schools and the California Assessment Program, so that much of the information about their relationship comes from informal observation and discussion.

Dr. Sue Bennett and Dr. Dale Carlson
The California Assessment Program
California Department of Education,
721 Capitol Mall
P.O. Box 944272
Sacramento CA 94244-2720
916-322-2200

Dr. Ruben Carriedo
Planning, Research, and Evaluation Division
San Diego City Schools
4100 Normal Street
San Diego CA 92103
619-293-8047

Dr. Grant Behnke and Robert Raines
Testing and Consolidated Applications
San Diego City Schools
4100 Normal Street
San Diego CA 92103
619-293-8509

Linda Carstens
Evaluation Department
San Diego City Schools
4100 Normal Street
San Diego CA 92103
619-293-8686

Dr. Shirley Weber and Susan Davis
San Diego City Schools School Board
4100 Normal Street
San Diego CA 92103
619-293-8562

APPENDIX B

GROUP GRADING

Group grading is a technique for increasing the objectivity and reliability of assessments scored by people as opposed to machines.

A group of graders (also "raters" or "scorers") assembles in teams around tables, each with a leader, usually a person experienced in the technique. The entire group, including all the members of all the teams, first has to agree on standards. So sample papers, usually selected by experienced readers to illustrate all the possible scores, are distributed for the entire group to read.

This activity is called "calibration" in some cases, "standard-setting" in others. It is the most important activity of the group grading process and may take as much as half the time allotted (usually measured in days) for scoring a large number of papers.

The teachers (in the case of grading the California Assessment Program writing assessments) first assign grades to the samples and then discuss what made them award a 6 or a 4 or a 1. The group must reach consensus so that these samples can be used as standards. It can be a strenuous process, but it is extremely valuable because it allows teachers to see clearly what they reward in writing and to reconcile their value systems with those of the larger group.

Different kinds of scales have been developed as group grading spread throughout the writing instruction community during the late 1970s and early 1980s. "Holistic" scoring awards a grade to a piece of writing based on its total impact on the reader; it does not analyze the writing into constituent parts. "Analytic" scoring does that, awarding separate scores for rhetorical effectiveness, use of arguments, tone, and so on, including correctness of grammar and usage. "Primary trait" scoring looks first for the expected features of a genre of writing, such as reporting, narration, a complaint letter, even a poem.

The California Assessment Program essays are graded according to a modified primary trait scoring system. Each essay has a rhetorical effectiveness score, which assesses its mastery of the type of writing attempted; a general feature score, rewarding ability to achieve coherence or an attractive style; and a conventions score, which assesses ability to use the grammatical and idiomatic structure of the language.

The calibration sessions in group grading can be used to establish a "rubric," (scoring guide) or to apply an already prepared one. There is considerable room for interpretation of key terms. What exactly constitutes evidence? What does "logic" and "coherence" mean in practice? Without thorough discussion, the standards can remain foggy and the grading will lose reliability.

After the calibration session, with the standards fresh in their minds, the graders begin reading. In the classic form of group grading, popularized by the National Writing Project in the late 1970s, two readers read each paper, without seeing the grade the other assigned it. If the grades agree within a specified tolerance, a final grade is assigned, usually the aggregate of the two grades. But if the first grader gave the paper a 4 and the second a 2 or a 1, then a third reader, usually the table

leader, reads it, again without knowing the grades, only that there is a discrepancy.

However, in the California Assessment Program writing assessments and Advanced Placement examinations administered by the Educational Testing Service only one reader reads each paper and other techniques are employed to maintain standards. There are frequent pauses for calibration sessions, usually before beginning in the morning and after lunch for large-scale readings that spread over a few days. Papers already scored are circulated back into the piles to see if they get the same scores again. Table leaders randomly select scored papers and read them to see if they agree with the scores assigned.

Most people who have been involved with group grading find the degree of reliability remarkable. There is of course the occasional reader who cannot accept the consensus standard. Such a person has to be neutralized by re-reading the papers--and by not being asked to return for further grading.

The group grading process is the basic means of scoring performance assessments. Some assessments use the entire process, including the open-ended mathematics questions of the California Assessment Program grade 12 assessment and the Arizona Student Assessment Program. Others, such as the New York State Elementary Science Program Evaluation Test Grade 4 Manipulative Skills Test use a reduced form, in which there is some discussion but a much greater degree of constraint on the judgment of the scorers; and the Longfellow Elementary School Art Assessment (case V in this series), which also allows for discussion among graders but constrains them to a detailed, explicit rubric, and only three possible scores for each answer.

APPENDIX C

Old Test

a) **What is the perimeter of an equilateral triangle with side 6 cm?**

o 6 cm
o 12 cm
o 18 cm
o 24 cm

b) **Which of the following is a multiple of 18?**

o 24
o 30
o 60
o 90

New Test

A cardboard piece shaped as an equilateral triangle with each side 6 cm long is rolled to the right a number of times. If the triangle stops so that the letter "T" is again in the upright position, which one of the following distances could it have rolled?

o 24 cm
o 30 cm
o 60 cm
o 90 cm

In part (a) of the old problem, the student simply needs to know the definitions of *equilateral triangle* and *perimeter* and how to multiply 6×3. In (b), what is required is the ability to recognize or calculate multiples of 18. Both are rote computations; neither requires *understanding* of concepts.

The new problem requires the student to analyze the situation and use spatial reasoning to see what is required, then make the same calculations as those in both of the old problems. Thus, understanding of concepts within several strands of mathematics is required to solve a situational problem.

Performance Testing in the California Assessment Program: An Update

The California Assessment Program (CAP) is the cornerstone of accountability for student achievement in California's public schools. CAP annually tests all public school students in grades 3, 6, 8, and 12. In the past seven years, CAP has made major revisions to tests and expanded statewide testing to additional grade levels and content areas. Breaking away from exclusive reliance on multiple-choice tests, CAP has begun introducing more authentic, performance-based assessment that supports curriculum reform by providing educators, the public, and the State Legislature with data that indicate more directly what students actually know, how well they can think, and what they can do.

CAP tests employ a technique called matrix sampling in which each student takes only a small portion of a much larger test. The tests are designed for program evaluation rather than to produce scores for individual students. They provide an extraordinarily powerful and reliable indication of the breadth and depth of student learning with a minimum of testing time.

In 1987, CAP introduced its first performance-based test, the grade eight writing assessment, in which students are asked to write an essay in response to a writing topic. This assessment was extended to grade 12 in 1988. Another change at grade 12 occurred in 1989 when open-ended mathematics items, for which students must construct their own solutions, were administered, scored, and reported. In 1988, CAP also initiated pilot portfolio projects in English-language arts and mathematics, in which students and teachers collect and evaluate an array of student work throughout the school year.

Several recent documents describe these new performance-based assessments in English-language arts and mathematics. *A Question of Thinking: A First Look at Students' Performance on Open-ended Questions in Mathematics* describes the development of the open-ended portion of the grade 12 test and its underlying rationale. It also provides sample items, scoring guides, and analyses of the assigned scores. The format of the writing assessment, its structure, goals, and procedures, is presented in a widely disseminated pamphlet entitled *Thinking Through Writing*. CAP is currently producing its *Guidelines for the Assessment Portfolio in Mathematics*, which provides practical guidance drawn from three years of experience with the ongoing portfolio pilot project cosponsored by CAP and the California Mathematics Project. The portfolio project in writing, cosponsored by CAP and the California Writing Project, is also described in detail in the spring 1990 issue of "The Quarterly of the National Writing Project and the Center for the Study of Writing."

English-Language Arts

CAP introduced statewide writing assessment at grade 8 in 1987 and at grade 12 in December 1988. Students write essays in response to one of a variety of prompts designed to elicit the full spectrum of thoughtful student writing at each of these grade levels. CAP has produced an extended report, *Writing Achievement of California Eighth Graders, A First Look*, which describes the development of the writing assessment, its underlying rationale, and provides sample essays, scoring guides, and a summary of the results of the first test administration.

The direct assessment of writing has had an extremely beneficial influence on local instructional programs as documented by a recent study by the National Center for the Study of Writing. Among the indicators of the impact of the grade 8 writing assessment:

assessment complements and enhances the sixth-grade open-ended questions and conceptual/thematic multiple-choice items that were also field tested in the spring of 1990.

History-Social Science

CAP is adding assessments of student achievement in the area of history-social science to its tests for grades 6 and 12 and revising the grade 8 test as well. These tests will mark the first major step towards assessments that are fully aligned to the curriculum envisioned in the *History-Social Science Framework*. They will allow students to demonstrate breadth of learning as well as their ability to clarify issues, recognize relationships, determine causes and effects, interpret evidence, and argue for a position. The new CAP tests will not emphasize isolated facts, but will assess a deeper, more thorough study of eras, events, and original documents in their full social and cultural contexts.

Assessment modes under consideration include written tests, portfolios of student work, and integrated performance tasks. Moving in this new direction, CAP has field tested open-ended questions in history-social science at grade 6 and essay, short answer, and integrated performance tasks at grade 12.

All of this new activity in the California Assessment Program indicates the dedication of the California Department of Education to the development of assessments that challenge students to create performances and products that really matter. Through these new, more authentic measures of student performance, the program encourages the ongoing enrichment of local instructional programs which are the heart of California's educational reform movement.

For more information about CAP, contact the California Assessment Program office at 916/322-2200.

Nov '98

NEW DIRECTIONS IN SCIENCE ASSESSMENT

For more information, contact:

Kathy Comfort (916) 327-2616
or
Holly Mc Laughlin (916) 322-2200

OVERVIEW

New Directions in Science Assessment

Highlights of the New Science Assessment

New Directions in CAP Science Assessment

Spring 1990

California is responding to the nationwide call for improved education in the sciences with its new science framework and new, richer, more thought-provoking forms of assessment. The American Association for the Advancement of Science recently published a report, *Project 2061—Science for all Americans* that addresses the critical need for the reform of education in science, mathematics, and technology. If we are to succeed in international economic competition, to advance technologically, and to learn more about ourselves and the world around us, education in science will play a central role. *Science for all Americans* includes a set of recommendations to improve science literacy in the United States. Those recommendations are reflected in the new Science Framework for California Public Schools, Grades K-12, and in the new directions for science assessment proposed by the California Assessment Program (CAP). The new science assessment will match the curriculum and model exemplary instruction. Students will be required to demonstrate knowledge of the concepts and processes of science, the ability to solve problems, and to engage in hands-on performance tasks. Students will be given an opportunity to discover and construct the important ideas of science through inquiry and investigation.

Development of Performance Tasks

Since 1988, the science teachers, supervisors, and university faculty comprising the CAP Science Assessment Advisory Committee have developed and piloted performance tasks focusing on the scientific processes embedded in the life, physical, and earth sciences. Performance tasks were field tested in spring 1990 for grade 6 and will also be field tested in fall 1990 for grade 12, moving toward implementation in 1991 and 1992.

CAP conducted the statewide field test of performance assessment at grade 6 in conjunction with the Science Unit of the California Department of Education, the California Science Implementation Network (CSIN), the California Science Project (CSP), Lawrence Livermore National Laboratory,

and the Lawrence Hall of Science. By focusing on the essential outcomes of learning recommended in the new science framework, the performance assessment complements and enhances the sixth grade open-ended questions and conceptual/thematic multiple-choice items that were also administered in the spring of 1990.

Approximately 1000 schools participated in this field test of performance assessment. All the teachers participating in the administration of the performance tasks were trained by CSIN staff developers. Approximately 980 sixth grade teachers throughout the state participated, and additional trainings were conducted at the district and site level.

Over 50,000 sixth grade students in California had the opportunity to demonstrate their understanding of important scientific processes and concepts by "doing" science. The testing format consisted of five "stations" with one task per station. Each task took approximately 10 minutes, and the students rotated through the stations, completing all five performance tasks in one testing session.

The students engaged in hands-on activities and recorded their observations and conclusions on five student response forms. They were asked to observe, classify, sort, detect patterns, infer, formulate hypotheses and interpret results. The five tasks challenged students to integrate their manipulative and thinking skills with their knowledge of the content of science.

The performance tasks field tested in 1990 asked students:

1) to build a circuit out of the materials provided, to predict the conductivity of various materials, to test their conductivity, and to record the results;

2) to create a classification system for a collection of leaves, and to explain the adjustments necessary when a "mystery leaf" is introduced into the group;

3) to perform a number of tests on a collection of rocks, to record the test results, and to classify them based on information provided;

4) to use the limited equipment provided to estimate and measure a particular volume of water;

5) and to perform a chemical test on samples of lake water to determine why fish are dying.

In each of these tasks, the students were asked to move beyond the hands-on activity and to apply what they learned to the understanding of complex natural phenomena.

Open-ended Science Questions

In spring 1989, open-ended science questions were administered to a total of 8,000 sixth grade students. These questions actively engage students in creating hypotheses, designing scientific investigations, and writing about social and ethical issues in science. Students respond to the questions in one of three ways: by interpreting and entering data on a chart; by drawing a picture to explain an answer; or by writing a short, analytical paragraph. Each open-ended question asks students to compose a written response which requires 10 to 15 minutes.

The Future of CAP Science Assessment

When implemented in spring 1991, the CAP grade 6 science test is expected to include open-ended items and performance tasks, as well as multiple-choice questions. In 1990 CAP is also developing Golden State Examinations (GSE) in chemistry and biology in collaboration with the San Diego County Office of Education, and with support from the Milken Family Foundation. These exams will also include multiple-choice items, open-ended questions, and performance tasks. High schools throughout the state will field test the GSEs in spring 1990, and both the GSE biology and chemistry examinations should be ready for statewide use in 1991.

SAN DIEGO CITY SCHOOLS
Planning, Research, and Evaluation Division
Testing and Consolidated Application Department

PROGRESS REPORT OF THE IMPLEMENTATION OF THE PROPOSED NEW TESTING SYSTEM

November 6, 1990

ISSUE/CONCERN

At the February 27th Board of Education meeting devoted to the topic of student achievement and assessment, a special presentation was made dealing with the development of a new testing system (NTS). A feasibility study was described which was to occur this spring. What are the results of the feasibility study? What proposals are there regarding norm-referenced testing?

BACKGROUND

There have been a variety of efforts dealing with the components of NTS during the 1989-90 school year*. NTS is the result of efforts to move towards an assessment system which will better meet the district's changing needs over the next decade. It seeks to reduce the over-reliance on conventional multiple-choice norm-referenced tests, while simultaneously presenting opportunities for meaningful alternative forms of assessment. The goals of NTS are:

1. To improve the quality of all of our tests
2. To reduce the amount of classroom instructional time lost to mandated testing
3. To provide a flexible testing system which will better meet our changing needs

This brief report focuses on the norm-referenced component, and recent efforts to develop curriculum framework-focused exams (CFFEs). The 1989-90 school year was a time of planning, developing, and field testing of a variety of assessment efforts. The curriculum specialists from Educational Services Division collaborated with district testing staff and Psychological Corporation's assessment experts. Feasibility studies were undertaken with abbreviated versions of a nationally known standardized norm-reference test series, as well as a set of custom-developed criterion-referenced tests which were administered to students in a matrix sampling approach.

More than 400 teachers and 12,000 students were involved in the feasibility studies of new abbreviated versions of the **Stanford Achievement Test**, Eighth Edition, or with the set of **CAP**-like criterion-referenced exams. The **Stanford Achievement Test** series has been established over the years as a reliable, valid, and useful achievement series. The various editions of the **Stanford**, published by the Psychological Corporation, have been used by millions of children, teachers, and administrators since their introduction in 1923. The feasibility testing of the abbreviated version, the **ASAT**, represented an initial effort of developing a brief version of the nationally renown **Stanford Achievement Test** series. The CFFE exams were collaboratively developed by testing experts from the Psychological Corporation working with curriculum and assessment personnel from the San Diego Unified School District. (Additional details of the feasibility study design may be found on pages 6 through 8 of the attached component summary report.)

*Included with this brief summary report is a document, **The New Testing System Component Progress Reports**, which provides details of the progress in each of the components of NTS (Attachment A).

The primary purposes of the feasibility study were to determine if the assessment concepts of NTS were functional and appealing to the school site staffs providing instruction, as well as district personnel involved in curriculum development. Teachers and site administrators reactions to the new forms of testing were sought via questionnaire and informal interviews. Curriculum staff from Educational Services Division were extensively involved in the development of the CFFE exams, as well as reviewing the objectives, content, and the manner of testing students via the **ASAT**.

Results of the Feasibility Studies

A very brief summary of results from the feasibility studies follows. The **ASAT** functioned as desired; providing valid estimates of student achievement, while consuming less instructional time than the **CTBS**. Statistical estimates of the "error of measurement" were obtained which were within the range of expectations (see Attachment B). Comments from teachers revealed areas of support and concern (e.g., the format of the Grade 2 mathematics subtests of the **ASAT** are much more auditorially oriented than the corresponding subtests of the **CTBS**). In addition, the district and the Psychological Corporation were able to complete an "equating study" to determine the equivalence of achievement between the **ASAT** and the **CTBS** by analyzing test scores of students who were administered both tests this spring (see Attachment C).

The curriculum framework focused exams encountered a few unanticipated outcomes. The most profound was that the second grade CFFE was inappropriate. Not only did the Grade 2 exam present a very awkward testing situation for teachers and students (using the matrix sample approach in primary grade classrooms), but it demonstrated the need to consider approaching primary grade assessment via other approaches (e.g., performance-based assessment activities as opposed to conventional multiple-choice testing).

At all three grades tested in the feasibility study (grades 2, 5 and 7), open-ended mathematics items were included with the multiple choice matrix sampling exams. The open-ended items met with mixed reviews (e.g., "My students are not used to such thought provoking problems." and "... I'm happy to see the emphases on thinking skills. ... There has been too much emphasis in the past on the tools and not on using the tools to think and solve problems. I've taught 30 years and am really excited by the new frameworks, especially science and math."). Teacher questionnaire responses to the open-ended math items were interesting, and almost contradictory. The responses indicated that the teachers perceived the math items as being interesting and challenging, but also frustrating and difficult to the students tested.

The relatively low percent of students correctly answering several items on the CFFEs met with mixed reviews from the curriculum specialists. Since the CFFEs are criterion-referenced, the specialists noted that there was ample room for improvement, as more attention is directed to these new and important California curriculum objectives. Simultaneously, there were some concerns that the test items, on occasion, did not adequately measure the concepts sought to be tested. Detailed data for the CFFEs have only recently become available. Much more analysis is needed. Attention will be focused on the Grade 7 exams. Efforts to review Grade 5 data have been put on hold, due to limited resources and the potential changing nature of **CAP**.

Proposal of Change to the Districtwide Testing Program

As presented at the February 27th board meeting, and as discussed at the September 4th board meeting, the mandated program for districtwide testing needs to be modified. The requirement of testing all students in the regular instruction program at grades 9 and 11 is non-productive. The conventional nationally-normed tests that are available for the senior high school grades measure basic skills, and not what is taught, or what should be taught and assessed at those grades. The instructional time lost, and the other resources consumed, is not commensurate with the

information obtained. Knowledgeable district staff in curriculum and assessment, as well as secondary teacher representation on the Test Advisory Committee, believe that such testing is counterproductive relative to higher level instructional efforts. Hence, regarding the administration of nationally-normed standardized achievement tests as mandated by SDUSD Board of Education, it is proposed to reduce the grades tested from 5, 7, 9, and 11 to grades 5 and 7. Persons seeking senior high normative information are encouraged to look at the statewide testing program which seeks to measure and encourage instruction in objectives from the California curriculum frameworks--which go well beyond "basic skills."

Regarding the curriculum framework focused exams (CFFEs), it is proposed that the Grade 7 exams be refined and administered in a districtwide pilot using a matrix sampling format this coming spring. Further development and administration of the CFFEs will depend upon the degree of acceptance from various audiences, the utility of the data forthcoming from that districtwide effort, and the allocation of additional resources to properly complete the development. If "successful," then appropriate CFFEs would be developed and administered districtwide at Grade 5, and at other non-CAP grades below Grade 9 at individual school's discretion and expense.

Attachment E addresses some commonly asked questions regarding the proposals above.

RECOMMENDATION

1. Support the above proposals regarding changes in districtwide testing grade levels and the districtwide pilot of CFFEs at grade 7.

2. Develop detailed informational materials dealing with the nature of the new tests, and provide inservice opportunities to site and central office staffs.

3. Provide the Board of Education, site and central office staffs, community members, and the media information about the new norm-referenced tests. Prepare these audiences for the anticipated shift in normative values previously described in public sessions (and as illustrated in Table 2 of Attachment D).

BUDGET IMPACT

Funds from district, externally-funded projects (which require testing at other grades), and integration efforts (for the court-identified minority-isolated schools) have been budgeted to cover the anticipated costs of the above proposal. Unless unusual circumstances arise, no additional funds should be needed for the 1990-91 school year. Further developmental efforts on CFFEs for the 1991-92 school year will be dependent upon additional resources for that school year.

Report Prepared by Grant Behnke
November, 1990.

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

IV. THE ADVANCED PLACEMENT STUDIO ART PORTFOLIO EVALUATION: A CASE STUDY OF NATIONAL PORTFOLIO ASSESSMENT

Section I. Introduction

Those who find it difficult to believe that portfolio assessment is possible, much less valid and reliable, would be surprised to know that a national example of portfolio assessment has been operating successfully since 1972. The College Board's Advanced Placement Studio Art Portfolio Evaluation asks students to send in a portfolio of their work and then gives each of three portions multiple scores, with a final score computed for the whole submission. The scores are assigned by a panel of judges drawn from college and high school art faculty. There are no essays, no answers to questions, no multiple-choice items.

We chose the Advanced Placement Studio Art Portfolio Evaluation as a case-study in order to find out how art production (as opposed to appreciation, criticism, or history) can be reliably scored, and how an examination which uses group grading exclusively has been accepted by the educational community.

Please note that the usual order of Sections IV, Impacts on Teachers, and Section V, Scoring has been reversed in this case study, so that Section IV concerns scoring and Section V concerns teachers. The rearrangement allows the discussion of scoring to follow logically from the experience of the assessment.

Section II. Purpose and Rationale

The purpose of the Advanced Placement Studio Art Portfolio Evaluation is to certify the achievement of high school students in studio art in comparison to that of college students. It is a "high stakes" assessment, that is, there are serious consequences for the students involved.

The Portfolio Evaluation is one of 29 Advanced Placement examinations offered in 15 academic disciplines in 1990 by the College Board. All the examinations are designed, administered, and scored by the Educational Testing Service (ETS). The Advanced Placement program was developed

the College Board as a cooperation between colleges and high schools to enable high school students to earn college credit for advanced work. The program began in 1955 and the examinations were administered for the first time in May 1956.

Colleges in the program will accept successful passing of the examinations for college-level credit. Some colleges will grant exemption from freshman-level courses, while others require students to take the introductory courses, but grant a certain number of elective credits. Students can reduce the number of courses required to graduate from college by taking college-level courses in high school. In some cases, students take as many as four Advanced Placement courses and reduce their college requirements (and in private colleges, their fees) by as much as an academic year. The occasional student has graduated with seven Advanced Placement courses as a result of beginning in grade 10.

The rationale for choosing a portfolio assessment follows directly from the nature of art production. It is impossible to imagine any other forms of assessment which would allow examiners to judge student achievement in studio art. The same principle applies here as in judgments of musical ability, or some kinds of athletic performance such as diving, gymnastics, or skating. There are no adequate proxies to represent achievement.

Section III. Experience to date

The Advanced Placement Program

When the Advanced Placement Program began in 1956, 1,229 students took 2,199 examinations; 104 schools and 130 colleges cooperated. By 1990, the number of students taking Advanced Placement examinations was 333,000 taking 490,000 examinations, and 42 percent of the secondary schools in the United States participated. The number of colleges accepting Advanced Placement credits, 2,537, was more than 19 times as many as in 1955.

The prestige of the Advanced Placement program is so high that the quality of a secondary school is often judged by its Advanced Placement offerings, despite the fact that the courses are college-level and hence outside the standards to be expected of high school performance. The existence of the Advanced Placement program could be seen as an implicit criticism of the standards high schools expect of their students, but it seems not to be interpreted thus. Principals point with pride to the number of Advanced Placement courses they offer, and high schools with aspirations promote their Advanced Placement programs as soon as they can get teachers trained in College Board courses and find students willing to endure the pressure of preparation.

The College Board arranges conferences for school personnel interested in Advanced Placement courses; offers summer courses taught by college faculty to train teachers; and provides consultants and materials to schools involved in the program. The conduct of the examinations is overseen and reviewed by development committees for each discipline, consisting of college and high school faculty.

All Advanced Placement examinations (with the exception of the Studio Art Portfolio Evaluation) consist of an objective multiple-choice section and a free-response section, in a ratio which varies from subject to subject but is usually in the 40:60 percent range. ETS scores the multiple-choice sections by machine, and then arranges for groups of judges (called "readers") to score the free-

response sections. The scores are reported to the student and to colleges on a 5-point scale, where 5 is the highest score and 1 the lowest.

Most colleges say they will grant credit for courses where the student has scored 3 or higher. However, in practice, colleges do not always grant exemption from introductory courses, and they may or may not grant elective credits. This issue is of course unconnected with the assessment as such, since it concerns the College Board and the higher education institutions.

Readers for Advanced Placement examinations

The free-response portions of Advanced Placement examinations are scored by panels of readers. The size of the panels varies in number from five readers for music to 376 for English literature in 1989. There is a Chief Reader for each examination, who is the ultimate arbiter of scores and who has the responsibility of determining the cutting points and raw score scales that convert the readers' scores into the Advanced Placement 5-point scale.

The Chief Reader is always a college faculty member, since the standards are those of college, not high school. Readers sit at tables with a table leader who acts as a sergeant in the army of readers, maintaining standards and reading papers that present scoring difficulties, and reporting up the line to the Chief Reader when necessary. The "readers" in the Studio Art Portfolio Evaluation do not sit at tables, (as the detailed description below will explain), but their process is essentially the same.

The papers are graded by first establishing "types" of each grade. A table of readers will discuss the properties of each model paper until they have internalized the standards they embody. Then each paper is read by a single reader. To maintain these standards, table leaders will hold group readings of sample papers two or three times during a day of reading, and will also do spot-checks for standards by choosing papers randomly.

The panels consist of about 40 percent high school and 60 percent college faculty members. High school teachers can apply to become readers after their students have been taking Advanced Placement examinations for three years, but they may not be chosen to participate for a number of years because the backlog of teachers wanting to read is so large. Panels have about 15 percent new members each year and no reader may serve for more than six years. The readings take place over a couple of weeks in early June, so that Advanced Placement results can be reported to colleges by the fisrt week in July so they can make placement decisions for students entering in September. The readings are held in colleges which can accommodate the large number of people involved, so the readers live in a college dormitory for six days. Although the readers are paid, the reading is hard work and requires concentrating on student work for seven and a half hours a day.

General information about the Studio Art Portfolio Evaluation

Students prepare for the Studio Art Portfolio Evaluation in high school courses which may be separate from other art courses in high schools with large enrollments, or subsections of art courses in smaller high schools. The Studio Art Portfolio Evaluation preparation is usually taught in grade 11 or 12.

Studio art was added to the Advanced Placement program in two stages--the general portfolio in 1972 and the drawing portfolio in 1980. It is one of two Advanced Placement examinations in art. The other is Advanced Placement Art History, which has the usual multiple-choice and free-

response answers and is not the subject of this case-study. Advanced Placement Studio Art Portfolio Evaluation is not called an examination but a "portfolio evaluation."

Students submit a portfolio consisting of three subportfolios for one of two evaluations: the drawing portfolio or general portfolio evaluation. In the drawing portfolio, there must be six original works no larger than 16" by 20"; as many as 20 slides in a transparent plastic pack on a single theme (concentration); from 14 to 20 slides in a plastic pack illustrating breadth--no more than three slides each from any one category. A general portfolio is much the same, although only four works are required in the original work portion; films and videotapes may be submitted in the concentration section; and the breadth categories are restricted to drawing, use of color, design, and sculpture, with eight slides required in drawing and four each in the other categories.

The three sections are sent to ETS in brown manila folders which will take works measuring up to 16" by 20". Their size produces a considerable logistical problem when there are 4,200 of them, as there were in 1990. The folders are carefully labeled and coded so that students' names are concealed. No-one touches the labels covering students' names until the entire reading is over.

At the end of the reading, the contents are replaced in the folders and they are sent back to the students, except for some exemplars which, with permission, are photographed for the conferences and workshops offered to interested schools.

Works of art present an obvious problem for reading or judging: how do you look at six drawings or four pictures to judge the quality section of the portfolio? Obviously they have to be displayed all at once, since the group as a whole, not the individual works, is to be evaluated. The answer is a huge room--a gymnasium is ideal--in which the original works are displayed on tables.

Before the time of tables in a gymnasium (and when portfolios were fewer), the portfolios were brought in stacks to the readers sitting around tables. The readers themselves took out the contents of each, reviewed them and put them back. The process was fraught with danger to the art work and to the reader's temper, as the art works slipped off tables on to the floor and the reader dropped papers trying to retrieve them. Then "floor-viewing" was tried: original works were displayed on the floor, in the absence of any other large space. The placement resulted in back-aches for everyone, especially the helpers who set out the works for the readers to view. The procedure now used involves displaying the contents of each portfolio on tables. They are arranged and removed by a small army of young people with strong arms and quick legs, who lay out the works with a space between them and with the brown manila folder adjacent, and then replace the works in the folder and stack the folders at the sides of the room.

The two other subsections, concentration and breadth, are judged by placing the slides on light tables. The concentration and breadth collections are viewed twice, with reconciliation if the grades are too far apart.

In the concentration section, raters look at the plastic packs propped against the light desks and then check what the student has written. Students choose their own concentration and must explain it on a standard form. They are asked to comment on the sources of their ideas, the influences on them, and the technical assistance they received while working on the project. Examples of concentrations from the 1990 general portfolio were a collection of slides of jewelry

designed in Japanese style; a series of copies of famous depictions of women; explorations of various media, such as drawing with a pen, using color entirely without drawing, or using texture as the design; and variations on a theme such as a discotheque, or the three kinds of Greek columns.

In the breadth collection of the general portfolio, a separate grade is assigned for each of the categories required. The breadth collection is intended to show how the students have stretched themselves beyond the medium or material with which they are characteristically successful. In the breadth collection of the general portfolio, students do not choose their own media, but must submit examples of drawing, color, design, and sculpture, all carefully prescribed. For example, the sculpture submission must consist of two views of two different sculptures.

Here perhaps more than in the other two sections it becomes obvious that there are multiple paths to a single grade. The best breadth collection would show some level of accomplishment in all four areas, but that is rare. One or two categories may be outstanding and the sculpture pedestrian. Or there may be a pervasive lack of balance in all areas. The qualities sought are necessarily on a high level of abstraction to accommodate the different media.

The Studio Art Portfolio Evaluation in 1990

In June 1990, a panel of 18 readers judged all these subsections, moving from general to drawing and from concentration to breadth throughout a day in order to maintain the freshness of their response.

The gymnasium at Trenton State College in Trenton NJ was filled with tables lined up across the floor with a central aisle. On each table, the six drawings or four pictures by each candidate were laid out with about two feet of space between each. Coded folders lay near them. The rows were divided into threes, right across the room, and three readers were assigned to each three rows.

The variety of submissions was astounding: in addition to paintings, photographs, drawings, collages, there were illustrated books and even paper constructions--anything in fact that would fit into the folders and survive the journey to Princeton and then to Trenton. Looking at the variety in the general examination at first makes judging seem impossible. Can a collection of four photographs be compared with four oils, or collages, or Sister Corita-like serigraphs?

The process of judging, packing, and replacing the works moved so efficiently that it sometimes looked like a wave motion. As soon as the three readers finished their three rows, the helpers repacked the original works into the appropriate folders and set out new ones. It took two days for 4,200 sets of original works to get on to the tables, be judged, and repacked.

There was an air of concentration, almost reverence, in the huge space. If anything had to be said, it was whispered. Most of the time there was only the sound of the ceiling fans and the soft tread of the readers as they moved along, paused, wrote, and moved a few more steps.

Section IV. Scoring

Each grading session begins with a standard-setting session. A number of portfolios are put before the assembled readers, roughly illustrating all the possible scores. These examples have been

chosen beforehand by the Chief Reader for the whole evaluation and the "table leader" for each section, quality, breadth, and concentration. Their selection is guided by their experience of first-year college work in art, since the Advanced Placement Studio Art course is designed to be equivalent to an introductory course in a college or university.

Anyone looking for a rubric, or criteria, or a list of features for the evaluation of studio art would be disappointed. There are no analytic scales or "primary traits" as there are in evaluating writing. This is what Walter Askin, professor of art at the California State University Los Angeles, and Chief Reader in the early 1980s wrote in Evaluating the Advanced Placement Portfolio in Studio Art, (Princeton, NJ: Advanced Placement Program, 1985) about standards for the quality section of the portfolio:

> Factors that are included in assessing quality imagination; freshness of conception and interpretation; mastery of concepts, composition, materials, and techniques; a distinct sense of order and form; evidence of a range of experience; and, finally, awareness of art-historical sources, including contemporary artists and art movements (p.25).

However, Askin continues:

> It is not expected that every student's portfolio will reflect all of these considerations to the same degree (p.25).

The major portion of the book is devoted to illustration in color of examples from student portfolios. This fact underlines the essentially experience-based nature of the evaluation.

The concentration section has a little more definition: "this section seeks to encourage the kind of search that is most comparable to the sort carried out by by an artists working independently in a studio" (p. 26). Students are expected to explore a theme or an approach in the concentration section, just as an artist would experiment. The breadth section also has more definition than the quality section: "the focus is on a broad spectrum of abilities normally represented in a first-year college course" (p.28).

Seeking a clearer understanding than these statements provide of the standards used for judging the portfolios, we interviewed Askin by telephone and tape-recorded his replies.

Askin said that many different factors are involved in making the evaluations. For example, there may be formal factors in relation to subject matter, but sometimes there is no subject matter so the formal elements have to carry the load. Essentially the judgment is made ad hoc on the basis of experience. The readers assess the most important factors, informally averaging the quality perceived in the individual pieces. Readers must choose not what they would like but what should be expected of college-level accomplishment.

"What you're really after is a mind at work, an interested, live, thinking being. You want to see engagement. Recognition of it comes from long experience and you intuit it."

Askin pointed out that this is also true in other subject matters. "There are more things that join us together than separate us, " Askin said. "You can make those judgments as accurately as you can

in mathematics or in writing or in any other subjects. These other subjects frequently have much more difficulty than we do in the visual arts in agreeing on standards....You get a sense for copied work, a sense when there's engagement, when inspiration, belief, direct involvement are present or absent."

The "types," the portfolios chosen to exemplify each grade, remain on display at the side of the room as standards for comparison. The readers assign scores from 1 through 4 separately to each part. Because three readers grade the quality section, and the concentration and breadth sections are read by two readers--and further, each portion of the breadth section for the general portfolio is scored separately--the scores are manipulated by computer to arrive at a raw score to which the three sections contribute equally. This score can range from 1 through the maximum possible score, which is somewhat below 100 for the drawing portfolio and above 100 for the general portfolios.

Scores are recorded by aides as each reader finishes a section. The aides note discrepant scores and forward the portfolio to the Chief Reader, who will reconcile the scores with his own judgment. If he thinks it appropriate, the Chief Reader will talk with a reader and use the "types" to reinforce the agreed standards, but this will be accomplished quietly and out of earshot of the main scoring, which is conducted in virtual silence.

If a reader should happen to recognize the work of a student, the Chief Reader is immediately summoned to judge the work, and the other reader passes on without a glance at the student's work or score.

To produce the scores on the Advanced Placement Program's 5 to 1 scale, it is necessary to establish "cut scores," the boundaries between 5 and 4, 4 and 3 and so on. The cut scores are the Chief Reader's responsibility, but there is a great deal of input from a long debriefing meeting of all the readers and from statistical information supplied by computer. This information consists of historical data about cut-scores for the past few years; the composite and raw scores for the present year's candidates; and tables showing the consequences of choosing certain cut-scores, in terms of percentages of students receiving 5, 4, 3, etc.

When the cut scores are established, there is a meeting of program representatives, College Board and ETS personnel to discuss them. If the Chief Reader can justify a scoring configuration that is quite different from the previous year's, his or her judgment will be respected, but it is rare that differences are wide. Since readers can serve for six years and continuity among personnel is assured by a regular turnover of a few readers, the judgments about cut scores are informed by collective experience.

The scores overall are roughly distributed in a bell curve, with the largest number receiving a 3, but with fewer 1 scores than 5s. This probably occurs because readers know that colleges do not usually accept either 2 or 1 scores, so a 2 will perform the same function as a 1 without being quite such a devastating judgment on a student's work.

Section V. The Impact on Teachers.

When we interviewed Walter Askin on the scoring of the Advanced Placement Studio Art Portfolio Evaluation, we also asked him whether art is an inborn talent or can be taught. Askin believes

that teachers make all the difference: "A bright mind can learn to paint and draw well," he said. "The best student artists I saw are probably now surgeons, because many students take all the Advanced Placement courses they can, including studio art, in order to cut a year out of college. Students flourish where there are good high school teachers."

The influence of teachers is so great that "sometimes I thought we were grading teachers as much as students." Some teachers bring out the best in their students; other teachers get good results, but their students are engaged strictly on a technical level. The teacher will assign failsafe assignments, which the student performs perfectly. But the thinking is the teacher's, not the student's, in that case, and the result is lifeless.

We talked to teachers and students in order to get a sense of how the examination exerted pressure on them. On the whole they were not much interested in talking about the way Advanced Placement Studio Art Portfolio Evaluation is graded. They were focused on the experience of preparing for the examination.

The three teachers we either met or watched all believe that art can be taught. Talent will add to a person's ability to judge what has to be done to improve a product, but talent also needs instruction. They share Walter Askin's belief that a bright mind can learn the techniques of art. They have a striking confidence in their ability to motivate, manage, and cajole adolescents and awaken them to their own potential. They enjoy an easy relationship with their students: they were capable of informality, even affection, but they did not indulge lack of discipline.

The three teachers all teach a technique they believe is essential to the development of artistic ability: they teach their students to criticize each other's work constructively. They make the class into a critical circle so that students will learn how to select works to go into their Advanced Placement portoflios. In all three cases, these criticism sessions are so well designed and the students learn to communicate the need for improvement so skillfully that the students are disappointed in the quality of criticism when they go to college or university. "Please teach these students how to critique," wrote one former student from college to Jack Bledsoe. One of Ray Campeau's former students complained bitterly about the quality of peer criticism in college courses: "they didn't know what to say." She missed the developed critical skills she had learned to depend on in high school classes.

Jack Bledsoe, who teaches Advanced Placement Studio Art Portfolio Evaluation at Walter Johnston High School in Montgomery County MD, has been teaching the course since the early 1970s (the general portfolio has been in place since 1972), when his high school began offering Advanced Placement courses. He teaches Advanced Placement Studio Art as part of his senior studio art course: 15 of the students are preparing Advanced Placement portfolios and the others are taking the regular course. Bledsoe is the author of a section of a booklet entitled Teacher's Guide to Courses in Advanced Placement Studio Art Portfolio Evaluation, to be published by The College Board in July 1991. He has applied to become a reader at the annual Portfolio Evaluation.

Bledsoe paces the students so that they do not run into a time-crunch during the last few weeks of preparation. They must have 12 slides done for their concentration or breadth section by Christmas. He also recommends that they save a few dollars a month towards the $65 fee. He insists that students choose their own concentration topic and explain it coherently: "That's when

they hate me the most. I won't do it for them. They whine and cry until they've got it figured out." Some of the concentrations his students have chosen include diving figures; architectural designs; shadows; and night scenes.

Bledsoe will not predict his students' scores and finds he can't do so. He is not present during the final assembly of the portfolios and their packing into the oversize folders for transport by United Parcel Service. He points out that other Advanced Placement teachers (like his wife Donna Bledsoe who reads Advanced Placement English examinations for ETS) can do nothing to help while their students are taking the examination. The final selection of the portfolio's contents is equivalent to the written examination, so he leaves it entirely to the students.

One student preparing for the Advanced Placement portfolio assessment in 1991 is a paraplegic. Bledsoe has cut down the scale of his productions from the usual 16 x 20 inches so that he can physically manage them.

Raymond Campeau, art teacher at Bozeman High School in Bozeman, Montana, came later into the Advanced Placement Studio Art Portfolio Evaluation program than Bledsoe, possibly because Bozeman, Montana, is not Montgomery County, Maryland. He said that his chief emotion on first entering students for the general portfolio evaluation in 1980 was fear--fear for two students who were being thrown into as what he and they thought of as an unknown competitive situation. The first two students earned a 4 and a 5, which made both Campeau and subsequent students confident. In the years since, Bozeman High School students have received only one 2, while all the other scores have been 3 or above. Between five and seven students submit portfolios every year, although a larger number take the class. Campeau became a reader for Advanced Placement Studio Art Portfolio Evaluation in 1989 and now spends a week in New Jersey judging student art every June.

Campeau believes that the Advanced Placement Studio Art Portfolio Evaluation course is a test of students' self-motivation: they must have the ability to envision a finished project and then work steadily towards it for eight or nine months, solving problems as they arise and timing production so that the entire portfolio--original works, concentration, and breadth--is ready at the deadline. Pieces have to be photographed for slides and final selections made for the collection of original works. Campeau discourages students who he thinks will not have the staying power, to avoid disappointment and frustration later in the process.

One of Campeau's students is now making his living as a graphic artist, and most of the others became art majors. One has changed to business administration from art. All cherished the experience of working for the Advanced Placement Studio Art Portfolio Evaluation examination as a vital contribution to their growing up.

Walt Bartman, art teacher at Walt Whitman High School in Montgomery County, MD, showed us how ideas are stimulated for concentrations. His students have the same dedication as those of Bledsoe and Campeau: five minutes before class began a few students were already looking at the slides to be shown that day, discussing their qualities, and comparing them to other work. In essence, they had started the class before the teacher had even arrived.

Bartman showed students the slides to help generate ideas for concentrations, which he, like

Bledsoe, regards as the real challenge of the Advanced Placement portfolio. The slides were primarily of best student works of previous years with a few masters randomly added. Bartman used the slides to sharpen the students' sense of connection between themselves and their work: "I need you to challenge yourself," he told the students. "Art is a mirror of yourself--people can see the knowledge in your paint....Look for the moment of emphasis in the painting and work on your awareness."

It is important, he thinks, for students to feel that they are doing something real for which they receive attention and critical feedback. Students in his class want to know specifically what they should improve. They prefer detailed criticism to a "Yeah, that's fine."

Like the students in Bledsoe's and Campeau's classes, these students refute the current impression that high school students don't care about their work and are generally lazy. When provided with something to care about for themselves, an appropriate environment, and a teacher they want to please, they will rise to the highest expectations set for them.

All three of our teachers also reported that their students are not routinely given credit for the freshman course which Advanced Placement Studio Art Portfolio Evaluation is intended to replace. Unless they put a great deal of energy into insisting on credit--and thus earning the ill-will of counselors and the art department--students all too frequently find themselves in beginning courses, drawing elementary objects. Since their year in Advanced Placement Studio Art Portfolio Evaluation has taken them far beyond this stage, they resent being treated as neophytes as well as having spent money and time without the expected reward. Colleges for their part defend their unwillingness to recognize the students' earned credit on the grounds that students need to be trained in "their" way.

A student interviewed for a Christian Science Monitor article, "The Art of Testing Young Artists," (1 November 1989), complained that she was being praised in college for works that she feels are not her best: "I think they should have exempted me from the entry courses...I began to slack off and my work just has not been what it could be."

Professor Theodore Zernich, Director of the School of Art and Design at the University of Illinois Urbana-Champaign, defends his university's decision not to give either course exemption or course credit on the grounds that development in art is qualitatively different from that in other academic fields. He believes that growth in art is recursive and that people "can always get better...There is no end-state for a first-year student in art." To be able judge whether object A is better or worse than object B, you must have an objective standard, which is lacking in a highly variable field like art. Although the ETS assesses portfolios by a time-honored process of consensus, judgment depends crucially on the quality of the judges. He sees so much growth at the end of the first year of a college art course that he cannot agree to granting exemption from it.

Without agreeing with either the students or the college faculty, we can say that the dissatisfaction with the current arrangements calls for the attention of the College Board to establish guidelines for both colleges and prospective students. It is not a question of compelling colleges to give credit, but making clear to both parties the expectations that each has of the other. Students go into the program expecting credit for college-level work done in high school, and colleges expect that students will appreciate the college's concern for their artistic development.

Section VI. Costs

Students will pay $65 for each Advanced Placement examination in spring 1991. $6 per student is returned to the school to cover its administrative costs or to help some students pay for the examinations. In addition, about $15 per student is available for students with financial need. The Advanced Placement program as a whole is now self-supporting, although for the first 30 years of its existence it was subsidized by the College Board.

However, the flat fee per student does not reflect the total cost to teachers and students involved in the Studio Art Portfolio Evaluation. It does not include the costs of the teachers' time spent coaching students, often outside normal working hours; or the costs of materials, often borne by the students themselves. A realistic picture of what the program costs (especially for comparison with similar programs, if any) would have to include these expenditures.

Section VII. Legal issues and impacts on minority populations

There has been no challenge to the Advanced Placement Studio Art Portfolio Evaluation related to minority populations. Although data on the ethnicity of students are collected by self-report (as they are for all Advanced Placement examinations), the breakdown for ethnic populations was not made until we requested it for this study. This demonstrates that ethnic differences have caused no problems for teachers, students, or the Educational Testing Service.

Seventy-three percent of students in both drawing and general portfolios are white. All minority are represented, but sometimes in small numbers: for example, only six students who identified themselves as American Indians entered drawing portfolios. The largest group of minorities participating were Asians and Asian Americans, who made up almost 10 percent of students submitting either drawing or general portfolios. These are all figures for students within the U.S., where the total submissions came to 4,105 for 1990; the numbers rise slightly if students from other countries (American schools abroad and in territories) are added.

Figures for the impact of minority status on scoring are not available for Advanced Placement Studio Art, because the standard model uses multiple-choice scores and has not yet been adapted for other kinds of scoring.

Alice Sims-Gunzenhauser, the ETS coordinator for Advanced Placement Studio Art Portfolio Evaluation, reported that she has responded to perhaps 20 telephone calls about the 1990 examination. Twenty calls out of 4,200 submissions is not a high rate of dissatisfaction. Some calls were from parents complaining that their children were penalized or disqualified because the teacher had not given students the right instructions. Sims-Gunzenhauser says that she suspects more than 20 scores are questioned by those who receive them, but only that number felt strongly enough in 1990 to write or call ETS.

All students taking Advanced Placement examinations are asked to complete a response form. Most question a single item, but cannot dispute results since the forms are filled out when the portfolio is submitted. The development committees read student responses in order to produce better examinations in subsequent years.

In unusual circumstances, an examination will be rescored, but this is extremely difficult with Advanced Placement Studio Art Portfolio Evaluation, because the portfolio would have to be recalled after it has been sent back, and there is a risk that it might be tampered with before resubmission.

Talking to teachers and students, whether they earned a 5, a 3, or a 2, revealed no dissatisfaction with the grading process. They accepted it as a given.

Section VIII. Discussion

Advanced Placement Studio Art Portfolio Evaluation is, as we pointed out, the only national portfolio examination--and the only portfolio examination known to us that attempts a composite grade for the whole portfolio, as opposed to grading the contents individually. If subjective grading is as big a threat as administrators and legislators say, then the grading of Advanced Placement Studio Art Portfolio Evaluation should be controversial. But it doesn't seem to be.

One obvious reason is that the nature of art is not conducive to "objective" evaluation. It is impossible to imagine a multiple-choice test for Advanced Placement Studio Art Portfolio Evaluation. For this examination, ETS has refined the jury method of judging works of art, where collective judgment substitutes for objectivity.

There is another somewhat chilling reason why there seems no opposition to Advanced Placement Studio Art Portfolio Evaluation's portfolio assessment. Art simply isn't considered important in American education. Jack Bledsoe, art teacher at Walter Johnson High School in Montgomery County, MD, referred to it in passing as "one of the minor subjects" in high school. Since the toughening of graduation standards in states and school districts, with their demands for more units in English, mathematics, science, and social studies, there is no time for electives and art has been squeezed out. It does not seem essential, and with the 1990s emphasizing workplace skills, its plight is likely to worsen. Because art is cosidered "minor," or a frill, it doesn't matter much how it is assessed.

However, this may not last for ever. The state of Maryland has made fine and performing arts a graduation requirement, putting a higher value on the arts than they had before. Of course, the other arts (music, dance, theatre) will also satisfy the requirement. But more attention may be paid to Advanced Placement Studio Art Portfolio Evaluation as a model for assessment, since what is required is usually evaluated as well.

If Maryland (and other states which may follow its enlightened lead) need a model of studio art assessment, the Advanced Placement Studio Art Portfolio Evaluation is ready and waiting. Its assessment manages to be faithful to the nature of the object assessed and at the same time psychometrically sound enough to survive 20 years of experience. It would also repay investigation by those who are sceptical about portfolio assessment and its effect on learning.

APPENDIX A

SITE VISITS AND CONTACTS

Ruth Mitchell visited the grading sessions at Trenton State College in Trenton NJ on 6 June; and interviewed students and an Advanced Placement Studio Art Portfolio Evaluation judge in Bozeman MT on 17 and 18 September. Kathleen Fulton of OTA, Amy Stempel of CBE, and Mitchell met with Alice Sims-Gunzenhauser of the Educational Testing Service (ETS) in Princeton NJ on 2 October. Stempel visited an Advanced Placement Studio Art class at Walt Whitman High School, Montgomery County, Maryland; Mitchell and Stempel interviewed Jack Bledsoe, Advanced Placement Studio Art teacher at Walter Johnston High School, Montgomery County MD; and Mitchell conducted telephone interviews with Walter Askin of the California State University, Los Angeles, and Theodore Zernich of the University of Illinois, Champaign-Urbana.

Walt McDonald, Director, Advanced Placement Program
Educational Testing Service
Princeton, NJ 08541
609-734-1457

Alice Sims-Gunzenhauser
Educational Testing Service
Princeton NJ 08541
609-734-1350

Raymond Campeau and students
419 South Grand Ave.,
Bozeman MT 59715
406-587-4392

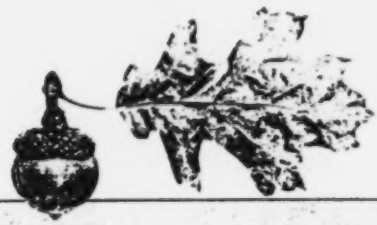
Walter Bartman and students
Walt Whitman High School
7200 Whittier Blvd.
Bethesda MD

John Bledsoe
Walter Johnson High School
Bethesda MD

Walter Askin
California State University, Los Angeles,
Los Angeles CA

Theodore Zernich, Director
School of Art and Design
University of Illinois
Urbana-Champaign IL

APPENDIX B



Advanced Placement Course Description

Art

Studio Art, History of Art

May 1991
The College Board

138

amination Grades

m *Guidelines on the Uses of d Data* (1988).

r placement and credit pur-

of particular tests through familiar with the content of

the performance levels and

ied by examination at a pre- est content meet their own

redit policies, making clear gain access to them, what it may be awarded for each granted meet degree require- ay be obtained.

ted by examination on stu-

varded by examinations that oviding this avenue to credit her unrelated costs.

oard test scores and related to the Chairman of the Ex- Regional Assembly of the f the particular use that is tter will be followed up and initiating the query or com-

AP Art: Studio Art and History of Art

In art, three Advanced Placement course descriptions and evaluations are offered: (1) Studio Art—General Portfolio, (2) Studio Art—Drawing Portfolio, and (3) the History of Art. The two Studio Art portfolios are designed for students who are seriously interested in the practical experience of art. These students who plan to attend universities or art schools that require a general art course at the freshman level should select the general portfolio offering. Those who plan to attend art schools or universities that require a year of drawing rather than or in addition to a general art course at the freshman level should select the drawing portfolio. Studio Art students are allowed to submit only one of the portfolio offerings in a given year; it is possible, however, that certain students may wish to submit one of each of the Studio Art offerings in successive years.

The History of Art course is designed to introduce students to the understanding and enjoyment of works of art. Students may present themselves as candidates in both Studio Art and History of Art. The descriptions of the two Studio Art portfolios appear immediately, followed by the description of the History of Art course.

Studio Art: General Portfolio and Drawing Portfolio

Advanced Placement Studio Art is not based on a written examination; instead, students submit portfolios for evaluation at the end of the school year. Since no standard, universally valid studio art course can or should exist, the Development Committee in Advanced Placement Studio Art has chosen to suggest guidelines focused toward the submission of an AP portfolio rather than to delineate a specific course. The portfolios are designed to allow maximum freedom in the structuring of AP Studio Art courses. Therefore, the major responsibility for creating an Advanced Placement course in art and preparing the work to submit for evaluation belongs to the participating teachers and students.

1

BEST COPY AVAILABLE

The Development Committee in Studio Art has had the counsel of both school and college teachers in defining the scope of work that would be equivalent to that of introductory college courses in studio art. Because art courses vary from college to college, the guidelines provided for AP Studio Art are not intended to describe the program of any particular institution, but they do reflect the coverage and level typical of good introductory college courses. Teachers are encouraged to give the course the shape that they feel their particular situation suggests. Therefore, the Development Committee in Advanced Placement Studio Art chooses to provide guidelines rather than a precise course description.

AP courses should reflect three major concerns that are constants in the teaching of art: (1) a sense of quality in a student's work; (2) the student's concentration on a particular visual interest or problem; and (3) the student's need for breadth of experience in the formal, technical, and expressive means of the artist. Advanced Placement work should reflect these three areas of concern: quality, concentration, and breadth.

The Advanced Placement Program in Studio Art is intended for highly motivated students who are seriously interested in the study of art. Most AP students will probably have had previous training in art, but such training is not a prerequisite for entering the program any more than it is for entering the introductory course in college. Students should, however, be made aware that Advanced Placement work involves significantly more time than the typical high school course and that the program is not for the casually interested.

The quest for quality of both production and experience in the Advanced Placement Program in Studio Art makes active demands not only on the students but also on the teachers and, indeed, on the school itself. Ideally, classes should be small enough so that teachers and students can work in close cooperation, extended time blocks should be allotted for instruction, and the teachers' other responsibilities should be reduced to reflect the greater demands of the program. However, since few situations are ideal, the course has been taught in many different ways: for example, as a separate, one-year class; as a separate program of study for AP students who meet during a general art class period;

BEST

140

as independent study for a few highly motivated students. Th introductory college course usually meets for three two-perio sessions a week, and such a schedule is preferable to the five one hour sessions a week typical of high school. Although the cours is designed as a one-year program, some schools may prefer t extend it over a two-year period. In such cases, the most recentl published AP Studio Art poster, detailing current requirement for each of the portfolios, should be consulted at the beginning o the second year of the course so that any changes in the evalua tion materials required can be taken into account well before th materials are to be submitted.

It is recognized that students will need to work outside th classroom, as well as in it, and beyond scheduled periods. Student should be considered responsible enough to leave the art room o school if an assignment requires them to do so, and homework such as maintaining a sketchbook or a journal, would seem to b a necessary component of instruction. Where museums and gal leries are accessible, teachers are encouraged to use them a extensions of school and to allot class time accordingly. In addi tion, art books, slides, and reproductions can provide importan examples for the serious study of art.

Basic Structure of the Portfolios: General Portfolio and Drawing Portfolio

147

In Studio Art, students present selected materials from the work they have done during the Advanced Placement course for eval uation at the end of the year by a group of artists and teachers The preparation of the portfolio for the Advanced Placement eval uation requires forethought to ensure that work submitted fo the various sections will be appropriate. The portfolio is a vehicle through which students can demonstrate their ability to deal with the fundamental concerns of the visual arts so that they have the possibility of earning credit or advanced placement, or both, on the college level.

The materials that the student is asked to present for evaluation fall into three sections. All three sections are required and carry equal weight, but students are not necessarily expected to

COPY AVAILABLE

perform at the same level in each section in order to receive a qualifying grade for Advanced Placement. The order in which the three sections are presented is in no way meant to suggest a curricular sequence; in fact, a good Advanced Placement Program in Studio Art would stress the three concerns—quality, concentration, and breadth—from the beginning of the course to its end. The works presented for evaluation may have been produced in art classes or on the student's own time, and may cover a period of time longer than a single school year.

Section A: Quality

"Quality" has generally been avoided as a criterion for evaluation in art education because it is difficult to identify and even more difficult to define. Nevertheless, it remains a basic concern in art.

By quality, the Development Committee in Studio Art refers to the entire work, involving all its aspects from concept to realization. Quality can be found in very simple works as well as those that are more elaborate. A student is asked to define through carefully selected examples of his or her own work his or her idea of "quality"—what is good work, *what succeeds in its own way.*

An awareness of quality does not come automatically but must be encouraged and developed. The importance of raising aesthetic questions in art classes cannot be overemphasized; by confronting such questions, students can more effectively involve themselves in art. One way in which the teacher can help develop an awareness of quality in students is by establishing a continuing critical dialogue with them, individually and in groups, about the works produced. Students should also be encouraged to give critiques of each other's work. Another means is to introduce the works of great artists, ancient and modern, into the discussion when they are appropriate to problems the class is considering. Care should be taken, however, to ensure that such works are used as departure points for the students' own conceptual and aesthetic inventions and critical judgments rather than as objects for them to imitate.

Although quality is an important component of work submitted for all sections of the evaluation, it can most readily be identified and evaluated in works presented in their original form. There-

fore, this section requires that students submit original works, four for the General Portfolio or six for the Drawing Portfolio.

Section B: Concentration

Students should be encouraged to explore a personal, central interest as intensively as possible. The Development Committee in Studio Art has therefore decided that one-third of the evaluation will be concerned with the student's presentation in depth of an aspect of his or her work in which the student has invested considerable time, effort, and thought. For example, a student might take a subject as general as color and attempt to investigate and experiment with it in a variety of ways—from paint to light. On the other hand, he or she may find a small but personally preoccupying visual symbol, or a material, or a process, or even a reference from psychology, politics, history, or philosophy that he or she can explore in depth by any available means. The project may even be a single work, such as a book to which the student has devoted a considerable amount of effort and thought over a period of months and which may be complete or still in the process of modification at the end of the course.

Whatever the content of the concentration, it should be a body of related works based on the student's interest in a particular visual concern. It should reveal the evolution of an idea and should focus on a process of investigation, discovery, and growth. A concentration is *not* a selection of a variety of works produced as solutions to class projects, or a selection of works with different intents. It should emerge from a plan of action and/or a predetermined line of investigation. It should not be construed as a detailed documentation of the long technical processes necessary for the completion of a simple work, such as the processes involved in castings, firings, and photographic and printing techniques. Group projects, teacher-assigned exercises and/or projects that merely require an extended period of time to complete should be avoided.

A written commentary must accompany the work. The required commentary provides the evaluators with basic information about the concentration. A form for the written commentary is provided to students in May when they are preparing the portfolio

materials for submission. Students will be expected to fill in the information requested on the form. An outline of the specific information requested on the Written Commentary Form appears on the Student Poster, copies of which are sent to the school to be distributed to students in the fall. The first instruction on that form asks the student for a brief definition of his or her concentration project. Students should be encouraged to think about how they will respond to that question early in the year, when they are first deciding the direction their concentration will take.

Section C: Breadth

Broad explorations are as important as concentration. Students should be encouraged to work in a rich variety of means and materials in order to extend their repertoire of visual experiences. Therefore, part of the student's work in Studio Art should show evidence of both intellectual and perceptual range. Thus, during the course the student should be introduced to a variety of problems in drawing (in the case of the Drawing Portfolio), or in drawing, color, design, and three-dimensional work (in the case of the General Portfolic). When appropriate, such investigations should encourage approaches and media that are new to the student.

Technical Specifications: General Portfolio and Drawing Portfolio

The following technical specifications are to be observed in preparing evaluation materials for submission to the examiners:

150 1. Students' names, or any other information that identifies them or their schools, should NOT appear on any of the materials; such information may remain, however, if its deletion would damage the work on which it appears. The Advanced Placement Coordinator at each school will receive instructions on how students should label their works and will give these instructions to the students.

2. Careful attention should be paid to the limitations on the size and the number of materials submitted for each section of

the evaluation. Since a standard portfolio for each student will be sent to the Coordinator, students should not plan to submit works in their own portfolios. Students should be advised NOT to exceed the maximum number of works, or to submit oversize works, or to send original work for sections that require slides. Extra works, oversize works, and originals sent in place of slides will not be evaluated.

3. After the students have received instructions for labeling and packaging their materials from the Coordinator, they should follow those instructions and must submit their prepared materials to the Coordinator. It is important that materials be submitted on time.

4. The slide packaging provided with the portfolios will accommodate only 20 slides for each section and only slides mounted in cardboard or plastic.

5. Tapes, videotapes, and films should be submitted in their containers. Films may be either 8–mm or super–8. Films and videotapes may not be submitted in Section A or Section C.

6. Students should be advised that, when photographing flat works of art, they should place the work against a neutral or nondistracting background and avoid covering the art work with a surface such as glass or acetate that will produce a glare. Advise students to photograph works as early in the year as possible to allow time to rephotograph, if necessary.

7. If a student wishes to show texture clearly in an art work in Section B, he or she should submit a detail slide of the work along with a slide showing the entire work.

8. The dimensions of the works photographed should appear on the cardboard mount of all slides submitted for Sections B and C. The medium of the work may be added if the student wishes.

9. Three-dimensional or fragile works should not be submitted in Section A.

10. All works and slides will be returned to the student in late summer. Each student is responsible for giving ETS the correct address for this purpose. Any portfolios sent back to ETS as undeliverable will be kept for six months and then destroyed. No responsibility can be assumed for their loss or damage.

11. Each student should have a copy of the Studio Art Poster that contains essential information from the course description

booklet. Schools that indicate they will present candidates for Studio Art evaluation will be sent a poster for each student. Additional copies may be obtained by filling out the AP Publications Order Form.

General Portfolio

Section A: Quality

In this section, students are asked to submit four original works in one or more of the media that can be executed on flat paper, cardboard, canvas board, or unstretched canvas. Such works may include drawings, paintings, graphics (including prints), photographs, diagrams, plans, animation cells, collage, montage, and so forth. These may be presented as four individual and separate works or as four items in a related series. The works for Section A, whatever form they take, will be evaluated on the basis of their overall quality.

Because of limitations imposed by shipping and handling the portfolios, work submitted for this section cannot be larger than 16″ × 20″, including matting or mounting. It is understood that students may have larger works of exceptional quality; these can be submitted in slide form in the other sections of the evaluation but may not be submitted here. Works that exceed the maximum size limits will not be evaluated. Works smaller than 8″ × 10″ should be mounted on 8″ × 10″ sheets. In order to protect the work, *all works on paper should be either mounted or matted;* materials that may become easily smeared should be protected with fixative or, preferably, with clear acetate. (Acetate wrapped around the work provides the best protection.) Works should NOT be rolled, glazed, or framed. The works may be titled, although titles are not required. Signatures should not appear on the works; a signature may remain, however, if its deletion would damage the work on which it appears.

Works submitted in Section A may be submitted in slide form in either Section B or Section C, but NOT both. Repetition among sections is not encouraged.

Section B: Concentration

The only works that can be submitted for this section in their original form are films or videotapes. Sculpture, painting, prints, graphic design projects, photography, architecture, etc., must be illustrated by color slides, even though the work itself may be black and white. Illustrated books or visual journals must also be documented and submitted in slide form. Environmental or conceptual projects should be described through slides and/or films, and through any written documents that are applicable. Films should NOT be used as a means of showing works of art such as drawings, paintings, or sculpture that can be photographed in slide form.

Whatever method of presentation is selected, certain limitations exist. A tape may be used with a slide presentation only if the tape is a cassette that electronically triggers the sequential showing of the slides. A film may be submitted with sound only if the film includes the sound track; a separate tape should not be submitted for the sound. Films and videotapes will be evaluated on the basis of visual rather than narrative elements. Films or slide presentations should not exceed five minutes in length; any presentation that does will be evaluated solely on the basis of its first five minutes. The maximum number of slides allowed is 20. For more detailed information, please see the section entitled "Technical Specifications" on pages 6–8.

Section C: Breadth

In this section of the presentation, students are given the opportunity to show the range of problems, ideas, media, and approaches they have undertaken during the course. A bibliography is provided on pages 15–18.

For Section C the student must submit 20 slides in four categories that are described below. Although it is possible that a single work might have qualities that make it eligible for more than one of the four categories, the student may NOT submit the same work in more than one category of Section C. Eighteen different works must be submitted. Works of photography may NOT be submitted in Section C.

153

Drawing

This category requires a student to submit slides of any eight drawings. The group of drawings as a whole should include works in which both line and tone are used. At least four of the drawings should show a variety of approaches to inventing spatial illusion; i.e., deep or shallow space using figures or objects in interior or exterior settings. The eight drawings as a group should demonstrate a range of expression as well as an exploratory use of materials.

Color (Two- or Three-Dimensional Works)

The color category requires each student to submit slides of four works in which color is the major component and which clearly demonstrate various means of color organization.

Examples:

1. Color theory
2. Color organization via a conceptual theory, such as primary, secondary, tertiary, or analagous color
3. Color used to indicate transparent effects
4. Symbolic color
5. Color used to create harmonic, lyrical, contrasting, or abrasive effects

Design (Two- or Three-Dimensional)

This category requires each student to submit slides of four works in which the principles of visual organization are the major concepts being displayed.

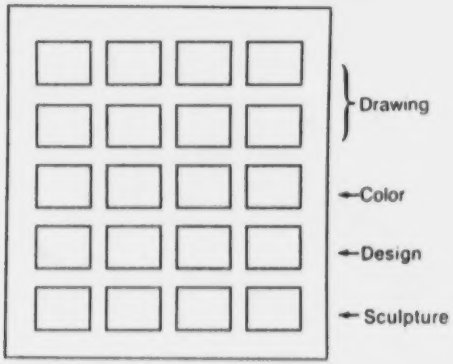
Examples:

1. Positive/negative space
2. Development of a repeat pattern or modular design
3. Typographic organization, layout, calligraphic exercises, three-dimensional logo design
4. Graphic designs for posters, book jackets, record jackets, etc.
5. Industrial design

154

Sculpture

This category requires a student to submit slides that show two different views of each of two sculptures constructed by using additive or subtractive techniques. Works submitted in this category may include figurative or nonfigurative sculpture, furniture, jewelry, containers, woven and metallic forms, and so forth. The slides for Section C should be arranged on the slide sheet as follows:



Slide Sheet

Drawing Portfolio

The existence of a drawing portfolio as an alternative to the general portfolio provides an option for both students and teachers of Advanced Placement Studio Art. The drawing portfolio allows for a more specific course of study that more readily parallels specialized drawing curriculums and programs in college and university art departments as well as in art schools. The Development Committee for Advanced Placement Studio Art, therefore, has based the drawing portfolio requirements upon a variety of basic drawing courses from major universities and art schools.

Drawing, whether employed as an end in itself or as a tool in the development of understanding form and content, is basic to most of the visual arts. Even though the artist uses drawing to express feelings, ideas, and observations of a personal nature,

there are certain technical skills and cultural theories and systems of "seeing" with which most student artists should be generally familiar. Being able to draw a representational image of a person or a piece of fruit in its situation of space and light in the three-dimensional world is an appropriate skill for the student artist to develop. Being able to draw from an "abstract" point of view or feeling, to express the "real" world abstractly through emotional or intellectual awareness is an equally valuable ability.

A fundamental drawing course should contain elements of both aspects. They are parts of each other; indeed it is hard to imagine how any work of art could exist in the form of "drawing" without displaying both representational and abstract (or "perceptual" and "conceptual") qualities. Perceptual skills relate to how we learn to see and represent objects and situations of our external world; that is, the development of the illusion of reality on a flat page. The artist wishing to do this might learn to "see" light and form and the surface and shape of things with acute observational perception. Conceptual skills relate to what our drawings *mean;* that is, what we *intend* them to communicate and what they eventually communicate to others. An exquisite perception of a tree before a darkening sky may transmit the concept of "beauty" to some persons or "loneliness" to others, since the content of a drawing is subject to interpretation.

In addition to perceptual and conceptual experiences, students should be encouraged to experiment with a variety of media in the hope that they will become familiar with tools and materials and thereby become aware of how a particular medium gives greater substance to a particular idea. No particular portfolio requirements are made regarding media and techniques. Works of photography, however, may not be submitted for any section of the drawing portfolio.

156 Finally, it should be noted that many drawings do not easily fit into any one category or definition of art; for this reason, the requirements for the drawing portfolio have been left as open as possible to allow both student and teacher the greatest freedom of choice. It follows that the student and teacher during the course of study should refer to many sources (see the bibliography on pages 15–18) as well as to original works whenever possible.

Drawing Portfolio: The Evaluation

Section A: Quality

In this section, students are asked to submit six original works (separate works or works in a related series). The works will be evaluated on their overall quality. Because of limitations imposed by shipping and handling the portfolios, work submitted for this section cannot be larger than 16" × 20", including matting or mounting. It is understood that students may have larger works of exceptional quality; these can be submitted in slide form in the other sections of the evaluation, but may not be submitted here. Works that exceed the maximum size limits will not be evaluated. Works smaller than 8" × 10" should be mounted on 8" × 10" sheets. In order to protect the work, *all works on paper should be either mounted or matted;* materials that may become easily smeared should be protected with fixative or, preferably, with clear acetate. (Acetate wrapped around the work provides the best protection.) Works should NOT be rolled, glazed, or framed. The works may be titled, although titles are not required. Signatures should not appear on the works; a signature may remain, however, if its deletion would damage the work on which it appears.

Works submitted in Section A may be submitted in slide form in either Section B or Section C, but NOT both. Repetition among sections is not encouraged.

Section B: Concentration

157

The drawing portfolio, like the general portfolio, requires the student to pursue a single concern in depth. While this section is open to a great variety of choice in the general portfolio, the drawing portfolio, by nature of the more specialized course, has greater limitations in regard to student choice. Some examples of projects that might be considered are: the investigation of a particular problem or subject through a variety of media, or conversely, the use of a single medium, such as charcoal, to explore a variety of problems; an animated film using drawing techniques; a mural based on a linear or drawing approach. Other suggestions are illustrations for literary works, or a series of

applications of ideas taken from master drawings. Another suggestion is a visual journal, not a casual sketchbook denoting daily recordings but one indicative of a sustained creative development of one or more ideas of personal interest. The choice of technique, medium, style, form, subject, and content are all decisions that are left to the student.

The works for this section should be submitted in the form of slides; a maximum number of 20 slides is allowed. Illustrated books or visual journals must also be documented and submitted in slide form. Films may not exceed 5 minutes in length. Films will be evaluated on the basis of visual rather than narrative elements.

Section C: Breadth

In Section C, the students are required to present evidence of their ability to work on a *wide variety* of drawing problems. The solutions to these problems should demonstrate that the student is able to pursue advanced drawing concepts from a broad base of exposure to, and experience in, drawing alternatives. Fourteen works are required in Section C; no more than three works focused on the same problem may be submitted. A minimum of 14 slides is required but a student may submit a maximum of 20 slides. The extra 6 slides may be used to show details of the work.

Some Suggestions for Section C

1. Evidence of training in use of perspective
2. Gesture drawings and/or contour drawings
3. Rendering of basic geometric forms in different spatial positions
4. Works dealing with positive–negative relationships
5. Psychological space (as in the work of De Chirico, Piranesi, or Matta)
6. Works in which line and/or tone are used to generate the illusion of three-dimensional forms
7. Integration of human figure with background
8. Drawings as social commentary or to express a particular attitude

9. Works involving collage and/or montage techniques or transfer
10. Works utilizing Cubist space: opening and closing space through interpenetration of form, shifting planes, etc.
11. Metamorphosis or transformation of an object or shape
12. A series of views of the same subject
13. Automatic drawing (as in the works of Jackson Pollock, Robert Motherwell, and Arshile Gorky)
14. Copies of exemplary works for particular ends
15. A section of an object (or objects) enlarged to a point of abstraction
16. Cross-contour surface drawings of crushed paper, cloth, the figure, etc.
17. Studies of wrapped objects
18. Drawings of dreams or of private fantasies
19. A study of an image divided into modules and then reassembled to create fresh relationships
20. Print or monoprint techniques

Media and Technique Suggestions for Section C Works

Students are encouraged to use a variety of media. Tools, materials, and approaches may include, but are not limited to, the following:

1. Pencil—black or colored, hard or soft
2. Conté crayon
3. Crayon
4. Pen and ink
5. Brush
6. Pastel or chalk
7. Charcoal
8. Ink or watercolor wash

159

Bibliography: Studio Art

The following bibliography is provided to serve as a resource for ideas and conceptual understanding. No single book or source on this list should be considered adequate to serve all interests or purposes. Selective reading and research should stand as one of the basic tools for the student's training and development.

The general portfolio consists of the three sections described below. All three sections are required.

SECTION A. QUALITY— FOUR ORIGINAL WORKS

The term quality, though elusive, can be used to express the success or the excellence of a work of art. Quality refers to the total work of art—the concept, the composition, and the realization of the artist's intentions. It can be found in very simple as well as elaborate works. For this section of the evaluation, you are asked to select examples of your own work in which these artists and teachers who evaluate them, with no other evidence, will recognize quality and will perceive that these works develop your intentions, both in concept and execution.

Quality can be seen more readily in original works than in slides. In this section, therefore, you are asked to submit four original works in one or more media. Your original works can be made on flat paper, cardboard, canvas board, or canvas that's been stretched. Such works include drawings, paintings, graphics, photographs, diagrams, plans, animation cells, collages, montages, and so forth. You may present four individual and separate works or four items as a related series. No three-dimensional works should be submitted in this section.

Because of limitations imposed by the shipping and handling of the portfolios, work submitted for this section may not be larger than 16" × 20", including matting or mounting. Students who have larger works of exceptional quality can submit them in slide form or other sections of the portfolio but not in Section A. Works smaller than 8" × 10" should be mounted on 8" × 10" sheets. In order to protect your work, all works on paper should be either mounted or matted; materials that may become easily smudged should be protected with fixative or, preferably, with clear acetate. Acetate may be wrapped around the work, providing the best protection. If you use an opaque overlay instead of acetate, fasten it to one edge of the work only, so that it can be lifted easily. If you present your work matted, a neutral color for the mat is usually advisable. Works should NOT be rolled, glazed, or framed. The works may be titled, although titles are not required. Signatures should not appear on the works; however, you should not risk damaging a work in order to remove a signature. A group of artists and teachers will evaluate the submitted works on their overall quality.

Works from Section A may appear in slide form in either Section B or Section C, but NOT both. Repetition of works among sections is not encouraged.

general portfolio 1989-90

SECTION B. CONCENTRATION— UP TO 20 SLIDES OR A FILM OR VIDEOTAPE

(Note: Illustrated books or journals must be photographed and submitted in slide form.)

In this section you are asked to show, in depth, your personal commitment to a specific visual idea or a particular mode of working. You are therefore asked to present an aspect of your work or a particular project in which you have invested considerable time, effort, and thought.

A concentration is a body of related works based on an individual's interest in a particular visual concern. It focuses on a process of investigation, growth, and discovery. It is not a selection of a variety of works produced as solutions to class projects, or a collection of works with differing intents. You are free to work with any idea in any medium that you like. In making your choice, however, remember that the concentration should grow out of a plan of action and/or a predetermined line of investigation. In this section, the evaluators are interested not only in the work you present, but also in evidence that you have thought out a particular project and have pursued it; the work should reveal the evolution of an idea or a process of investigation. Most concentrations involve a group of works. Whether your concentration includes a relatively small number of works or a greater number, the works should be united by an underlying idea that has visual coherence. Your concentration could consist of a group of independent works that share a single theme – for example, a variety of ways of handling what for you is a particularly interesting subject, or an in-depth study of a particular visual problem. Some concentrations involve sequential works—for example, series of studies that lead up to, and are followed by, more finished works.

A concentration should not be construed as a detailed documentation of the long technical processes necessary for the completion of a single work, such as the procedures involved in casting, firing, and photographic and printing techniques. Group projects, collaborations, teacher-assigned exercises and/or projects that merely require an extended period of time to complete should be avoided.

A written commentary must accompany the work in this section. The commentary form enables you to describe for the evaluators the plan of action or line of investigation that determined the focus of your concentration project. You are encouraged to work out the response to the first question early in the year, as you are defining the direction your concentration will take.

The Written Commentary Form asks for the following information:

1. Briefly define the nature of your concentration project.
2. What were the sources of the ideas represented in your projects? What influences (things you have seen, heard, read, felt, or imagined) affected your work?
3. List any technical assistance you received or resources (if any) you used while working on the concentration.

A form for your written commentary will be provided to you in May when you are preparing the portfolio materials for submission. You will be expected to fill in the information requested on the form.

The only works that can be submitted for the section in their original forms are films or videotapes. Sculpture, painting, graphics, photography, and architecture must be illustrated by 35mm color slides, even though the work itself may be black and white. Environmental or conceptual projects should be described through slides, films, and/or appropriate written documents. Films should not be used as a means of showing works of art such as drawings, paintings, or sculpture that can be photographed as slide form.

Whatever method of presentation is selected, certain limitations exist. Tapes may be used with slide presentations only if the tape is a cassette that electronically triggers the sequential showing of the slides. A film may be submitted with sound only if the film includes the soundtrack. A separate tape should not be submitted for the sound films and videotapes will be evaluated on the basis of their visual rather than narrative, qualities. Tapes and films may not exceed 5 minutes, the maximum number of slides allowed is 20.

SECTION C. BREADTH— 20 SLIDES

The concern for breadth is an attempt to avoid the over-refined and past patterns of success. The evaluators will be interested in your exploration of the rich variety of concern and materials that extend your intellectual and perceptual experiences.

In this section, you are able to show the range of problems, ideas, media, and approaches you have undertaken. Section C consists of the four categories described below; you are required to submit 5 slides of works for all of Section C. Works of photography may not be submitted in this section.

Drawing

For this category you are required to submit slides of one eight drawings. The group of drawings as a whole should include works in which both line and tone are used. At least four of the drawings should show a variety of approaches to investing spatial illusion or depth of shallow space using figures or objects in interior or exterior settings. The eight drawings as a group should demonstrate a range of expression as well as an exploratory use of materials.

Color

The color category requires slides of four works in which color is the major component and which clearly demonstrate various means of color organization. These works may be either two-dimensional or three-dimensional.

Examples:

1. Color theory, such as that explored in Impressionism, Pointillism or Fauvism
2. Color organization via conceptual theory, for example, primary, secondary, tertiary, or analogous color
3. Color used to indicate transparent effects
4. Symbolic color
5. Color used to create harmonic, tonal, contrasting, or discordant effects
6. Color used to create or intensify the expression of a particular emotion

Design (Two- or Three-Dimensional)

This category requires you to submit slides of four works in which the principles of visual organization are the major concepts being displayed.

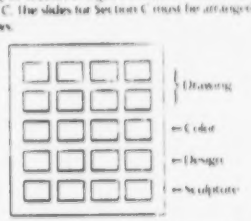
Examples:

1. Positive-negative space
2. Development of a repeat pattern or modular design
3. Typographic organization, layout, a calligraphic exercise, or three-dimensional logo
4. Graphic designs for posters, book jackets, record jackets, etc.
5. Industrial design

Sculpture

For this category, you must submit slides that show two different views of each of two sculptures created using additive or subtractive techniques. Works submitted in this category may include figurative or nonfigurative sculpture, furniture, jewelry, ceramics, assemblages, and metallic forms, and so forth.

For Section C, you must submit 20 slides. Although it is possible that a single work might have qualities that make it eligible for more than one of the categories, you may not submit the same work to more than one category of Section C. Eighteen different works must be submitted in Section C. The slides for Section C must be arranged in the slide sheet as follows:



Each slide should be labeled on the slide mount with the identification of the work and the category in which it belongs. You may abbreviate the titles of the categories — for example: D for Drawing, C for Color, Dn for Design, and S for Sculpture.

Bibliography

See the AP Course Description for a reference bibliography.

Advanced Placement Program
THE COLLEGE BOARD

See other side.

19002 01845 • H89830 • 271471 • Printed in U.S.A.

BEST COPY AVAILABLE

Note: In preparing your materials for evaluation, follow these instructions carefully:

1. Do not place your name or any other information that identifies you or your school on any of the materials. The AP Coordinator at your school receives instructions by May 1 on how to label your works.
2. Because a standard portfolio will be sent to the coordinator for you, do not plan to submit works in your own portfolio. Pay attention to the limitations given on this poster about size and number of materials that may be submitted.
3. Do not exceed the maximum number of works or slides required for each section of the portfolio. Do not send original work for sections that require slides, or send original works that are larger than 16" x 20". Extra works, originals sent in place of slides, or oversized works will not be evaluated.
4. For any slides submitted in parts B and C, the dimensions of the original works shown on the slides should be written on the slide frames. Titles may be included, if you wish. All slides should be cardboard or plastic mounted, not glass mounted. Films should be submitted in their containers and may be either 8mm or super-8.
5. When photographing works of art to be submitted in slide form, place the works against a neutral or nondistracting background, mask the art work rather than the slide, and avoid covering the art work with a surface such as glass or acetate that will produce a glare. Try to photograph your work early enough to make a second try if necessary.
6. If you wish to show texture clearly in an art work, submit a detail slide of the work along with a slide of the entire work.
7. Three-dimensional works and/or works easily broken (e.g., ceramics) and valuable metals (e.g., jewelry) should not be submitted in their original forms. The slide sections of the portfolio are meant to accommodate such works. The Advanced Placement Program cannot take any responsibility for damage and/or loss during the course of the grading if such works are included in the portfolios.
8. Retain duplicates of slides and papers. All materials submitted for the evaluation will be returned late in the summer directly to the address you indicate at the time of submission. No responsibility can be assumed, however, for loss of or damage to the materials.
9. Follow your coordinator's instructions for labeling and packaging your materials and submit your prepared materials to the coordinator sometime within the first two weeks of May. The exam fee for either portfolio is $62. It is important that materials be submitted on time.

general portfolio 1989-90

Note: In preparing your materials for evaluation, follow these instructions carefully:

1. Do not place your name or any other information that identifies you or your school on any of the materials. The AP Coordinator at your school receives instructions by May 1 on how to label your works.
2. Because a standard portfolio will be sent to the coordinator for you, do not plan to submit works on your own portfolio. Pay attention to the limitations given on the poster about size and number of materials that may be submitted.
3. Do not exceed the maximum number of works or slides required for each section of the portfolio. Do not send original work for sections that require slides, or send original works larger than 18" x 20". Extra works, originals sent in place of slides, and oversized works will not be evaluated.
4. For any slides submitted in parts B and C, the dimensions of the original works shown on the slides should be written on the slide frames. Titles may be included, if you wish. All slides should be cardboard or plastic mounted; not glass mounted. Films should be submitted in their containers and may be either 8mm or super-8.
5. When photographing works of art to be submitted in slide form, place the works against a neutral or nondistracting background; work with a surface such as glass or acetate that will produce a glare. Try to photograph your work early enough to make a second try if necessary.
6. If you wish to show bodies of work or art work, submit a color slide of the work along with a slide of the entire work.
7. Retain duplicates of slides and papers. All materials submitted for the evaluation will be returned late in the summer directly to the address you indicate at the time of submission. No responsibility can be assumed, however, for loss of or damage to the materials.
8. Follow your coordinator's instructions for labeling and packaging your materials and submit your prepared materials to the coordinator sometime within the first two weeks of May. The exam for other portfolios in May. It is important that materials be submitted on time.

The drawing portfolio consists of the three sections described below. All three sections are required.

SECTION A. QUALITY— SIX ORIGINAL WORKS

The best works, through choice, can be used to express the success or the excellence of a work of art. Quality refers to the total work of art — the concept, the composition, and the realization of the artist's intentions. It can be found in very simple as well as elaborate works. For this section of the evaluation, you are asked to select examples of your own work in which these artistic and technical qualities are present. You, rather than others who evaluate them, with no other evidence, will recognize quality and will perceive that these works develop your intentions, both in concept and execution.

Quality can be seen more readily in original works than in slides. In this section, therefore, you are asked to submit six original works. Separate works or works in a related series in one or more media. Your original works can be made on flat paper, cardboard, canvas board, or unstretched canvas.

Because of limitations imposed by the shipping and handling of the portfolios, work submitted for this section may not be larger than 18" x 20" including matting or mounting. Students who have larger works of exceptional quality can submit them as slide form in other sections of the portfolio but not in Section A. Works smaller than 8" x 10" should be mounted on 8" x 10" sheets in order to protect your work. All works on paper should be either mounted or matted; materials that may become easily smudged should be protected with acetate or, preferably, with clear acetate. Acetate may be wrapped around the work, providing the best protection. If you use an opaque overlay instead of acetate, fasten it to one edge of the work only, so that it can be lifted easily. If you present your work matted, a neutral color for the mat is usually advisable. Works should not be ruled, glazed, or framed. The works may be titled, although titles are not required. Signatures should not appear on the works; however, you should not risk damaging your work in order to remove a signature. A group of artists and teachers will evaluate the submitted works on their overall quality.

Works from Section A may appear in slide form in either Section B or Section C, but NOT both. Repetition of works among sections is not encouraged.

SECTION B. CONCENTRATION— UP TO 20 SLIDES OR A FILM

(Note: Illustrated books or journals must be photographed and submitted in slide form.)

The drawing portfolio, like the general portfolio, requires you to pursue a single concern in depth. While the section in the general portfolio is open to a variety of choices, the drawing portfolio, by nature of the more specialized course, has greater limitations. You must select a project on which you have invested considerable time, effort, and thought.

A concentration is a body of related works based on an individual's interest in a particular visual concern. It focuses on a process of investigation, growth, and discovery. It is not a selection of a variety of works produced as solutions to class projects, or a collection of works with differing intents. You are free to work with any idea or any medium that you like in making your choice, however, remember that the concentration should grow out of a plan of action and/or a predetermined line of investigation. In this section the evaluators are interested not only in the work you present, but also in evidence that you have thought out a particular project and have pursued it; the work should reveal the evolution of an idea or a process of investigation. Most concentrations involve a group of works. Whether your concentration includes a relatively small number of works or a greater number, the works should be united by an underlying idea that has visual coherence. Your concentration could consist of a group of independent works that share a single theme (for example, a variety of ways of handling what for you is a particularly interesting subject, or an in-depth study of a particular visual problem. Some concentrations involve sequential works — for example, series of studies that lead up to, and are followed by, more resolved works.

Some examples of projects that might be considered are: a particular problem or subject that has been investigated through a variety of media, or, conversely, through a single medium such as charcoal which is used to explore a variety of problems; an animated film using drawing techniques; a mural based on a linear or drawing approach. Other suggestions are illustrations for literary works, a series of appropriations of ideas taken from master drawings, or a visual journal. The visual journal should not be a casual sketchbook devoting daily recordings, but one that indicates sustained creative development of one or several ideas of personal interest. The choice of technique, medium, style, and content are all decisions left to the student.

Group projects, collaborations, teacher-assigned exercises, and/or projects that merely require an extended period of time to complete should be avoided.

A written commentary must accompany the work in this section. The commentary form enables you to refer to the evaluators the plan of action or line of investigation that determined the focus of your concentration project. You are encouraged to work out the response to the first question early in the year, as you are defining the direction your concentration will take.

The Written Commentary Form asks for the following information:

1. Briefly define your concentration project.
2. What were the sources of the ideas represented in your project? What influences or things you have seen, heard, read, felt, or thought about affected your work?
3. List any technical discoveries you recovered or inventions you used to any) while working on the concentration.

A form for your written commentary will be provided to you in May when you are preparing the portfolio materials for submission. You will be expected to fill in the information requested on the form.

The works for this section should be submitted on the form of slides mounted as cardboard; a maximum of 20 slides is allowed. Films submitted may not exceed 5 minutes in length. Films will be evaluated on the basis of their visual, rather than their narrative, qualities.

SECTION C. BREADTH — 14 to 20 SLIDES

In Section C, you are required to present evidence of your ability to work in a wide variety of drawing problems. The solutions to these problems should demonstrate that you are able to pursue advanced drawing concepts from a broad base of exposure to and experience in drawing alternatives. Fourteen additional works are required in Section C, no more than three from any of the categories listed below or from other categories of your own choosing or invention.

Suggestions for Section C Works

1. Evidence of drawing in one or perspective
2. Gesture drawings and or contour drawings
3. Rendering of basic geometric forms in different spatial positions
4. Works dealing with positive-negative relationships
5. Psychological space (as in the works of de Chirico, Picasso, Matta
6. Works in which line and or tone are used to generate the illusion of three-dimensional forms
7. Integration of human figure with still life or other background
8. Drawings in various communities or to express a particular attitude
9. Works involving collage, montage techniques or transfer
10. Works utilizing found space, layering and viewing space through interpenetration of hole-defining planes, etc.
11. Metamorphosis of [illegible] and light or shape
12. A series or sense of the same subject
13. Automatic drawing, intuitive works (as in Paul Klee, Kandinsky, Miro and Andre Gides
14. Copies or exemplary [illegible]
15. A section of an object [illegible] enlarged to a point of abstraction
16. Cross-contour surface drawings of crushed paper cloth, the figure, etc.
17. Studies of wrapped objects
18. Drawings of dreams or of private fantasies
19. A study in which an image is divided into modules and is reconstructed to create a fresh sense of relationships
20. Print or transfer techniques

Media and Technique Suggestions for Section C Works

1. Pencil, [illegible]
2. Conte crayon
3. Crayon
4. Pen and ink
5. Brush
6. Pastel or chalk
7. Charcoal
8. Ink or watercolor wash

General Directions

For Section C, you must submit a minimum of 14 slides. Because there are 20 slide spaces on the slide sheet provided, you may also submit extra slides to show details of up to but the 14 works if you choose. A maximum of 20 slides may be submitted.

Although it is possible that a single work might have qualities that make it eligible for more than one of the categories listed, submit the same work in more than one category in Section C.

Bibliography

See the AP Course Description for the bibliography.

drawing portfolio 1989-90

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

V. LONGFELLOW ELEMENTARY SCHOOL, BOZEMAN MT:

A SCHOOL-BASED ART ASSESSMENT

Section I. Introduction

A characteristic of performance assessments is that they can be devised locally and can yield sufficient information for a circumscribed purpose. They do not need the specialized knowledge and sophistication of nationally norm-referenced assessments.

An elementary school in Bozeman MT has assessed the effectiveness of an instructional program in art, using pre-post questionnaires administered by volunteer parents, and providing just enough information to estimate the success of the program. A local high school teacher, Ray Campeau, devised the assessment, which was carried out for a tiny outlay of money (although some costs are concealed). We focused on the program for a case-study because it is an example of a homegrown assessment of a subject usually neglected--learning about art.

It interested us more for its potential as a means of evaluating progress in learning about art and for involving the community than for its implementation, which was flawed by naivety about the technical aspects of assessment.

Bozeman MT is a town of about 22,000 inhabitants about 50 miles North of Yellowstone National Park. The town has gracious Victorian houses, some built by Southerners who migrated to Bozeman during or after the Civil War; some of the houses have now become fraternities for students at Montana State University. Its school district consists of a high school, junior high school, and five elementary schools (all named after American authors). The population is predominantly white middle-class, with some Native Americans and a handful of African Americans. Montana State University is a major employer, and there are some small manufacturing companies. There is some employment in the agriculture which surrounds the town and in the tourist industry stemming from the proximity of Yellowstone. Some high school students are bussed from a number of small outlying communities.
Bozeman has a tradition as an artistically active community, beginning at the town's airport, where metal sculptures of Canada geese fly over the arriving traveler, suspended over the staircase and

lobby on wires of varying lengths. It quickly became apparent that Ray Campeau is Mr. Art in Bozeman. He is a dynamic personality overflowing with energy and ideas about teaching art. He has owned art galleries in the town, currently operates a bronze art foundry, and heads a program at the high school which makes art a living presence in every part of the school. The high school library, a magnificently designed multi-level room, has colored-glass panels in its 20-foot high windows donated by a former student of Campeau. Stuffed paper figures hang over the stairs between the levels. An enormously oversize sculptured head sits on the floor by the book stacks. These are only the most immediately noticeable of a myriad of pieces of every size, shape, and material that make the library a three dimensional find-the-art puzzle.

The high school's corridors are the gallery for its art collection. The collection belongs to the student Art Club, which cares for and adds to the collection annually. Works by several well-known artists (Kathe Kollwitz, Alexander Calder, Mark Toby, Thomas Hart Benton) are included in the display of lithographs, paintings, and designs. Each year Campeau's students go on an art trip, just as high school bands go on tour or history students take trips to Washington, DC. Campeau packs his Art Club members into a bus (which also serves as a traveling hotel, in a pinch) for a tour of art galleries in the Western states. They always return with at least one piece to add to the student art collection.

In addition to Campeau's local activities, he has been a judge for the College Board's Advanced Placement Studio Art Portfolio Evaluation for two years; has been a member of the Educational Advisory Committee of the National Endowment for the Arts; is a teaching member of the National Arts Education Research Center in New York; and has recently been appointed to the Early Adolescent through Young Adulthood/Art Standards Committee of the National Board for Professional Teachings Standards (NBPTS).

Longfellow Elementary School is a K-4 elementary school with about 360 students. The principal is Michael Meagher, and there are thirteen regular classroom teachers, (all women), two librarians, and some specialist and categorical program staff. It has the support of a corps of mothers who have part-time jobs as aides in some classrooms, and also volunteer their time to assist instruction. Because of the town's isolation, there may be more women available to assist in the school than would be the case in larger cities, where the same women might be employed in fulltime jobs.

Section II. Purpose and rationale

The Longfellow Elementary School Art Assessment was designed as to evaluate the success of a program of art instruction, Art in Action. It is a program assessment, by no means "high stakes." Success in terms of the instructional program means the ability of children to use the vocabulary of art appreciation as they have been taught in the Art in Action curriculum.

The rationale for choosing the particular form of the assessment stems from the purpose in the context of the school, the district, and Bozeman itself.

The assessment is essentially outside the organizational structure of Longfellow Elementary School. Ray Campeau is not in a leadership position in Longfellow School, although he enjoys considerable prestige in the Bozeman district as a whole. Because he is not an official leader, he cannot use Longfellow School's resources freely in the administration of the assessment. He is dependent on volunteer help, which in our experience affected the design and the outcome of the assessment.

Longfellow, unlike some other elementary schools in Bozeman, might be described as traditional: its classrooms contain neat rows of desks, instruction is teacher-centered, and the school takes great pride in its success on nationally normed, multiple-choice tests. Ironically, the art assessment grew out of administrative caution about a new instructional program, begun in the school year 1989-90.

In 1988, Longfellow's scores on the Stanford Achievement Tests placed it at or above the Bozeman School District's scores in every test but mathematics computation. The school participates in the Onward to Excellence (OTE) program designed and disseminated by the North West Regional Laboratory (NWRL). The OTE program focusses on an academic area each year, with goal-setting and strategies for achieving improvement. In 1989, Meagher told us, Longfellow's staff decided that art should be the OTE focus, since academic success was so high already. (In 1990-91, the goal is improvement in recreational reading, to be assessed by charts kept at home and in school.)

The decision to focus on learning about art (not art production, which is a normal feature of the Longfellow elementary curriculum) coincided with other events in Montana art education. In 1988, the Montana Art Education Association and the Montana Office of Public Instruction, together with a consortium of seven school districts, applied to become one of the Getty sites for its discipline-based art education program. They were unsuccessful. So the Montana Art Education Association and the Office of Public Instruction developed their own summer institute modeled on a Getty institute. Teachers who attended the institute surveyed published curricula and school district curricula in art education. Teachers from Bozeman were given scholarships to attend the institute, and as result of their experiences with art education curricula, they made a recommendation to use Art in Action from Coronado Publishers in San Diego in the Bozeman school district.

All the Bozeman elementary schools adopted the curriculum, which was endorsed by the school board. Above the cost of the teachers' Art in Action instructional manuals and packages of large reproductions, about $3,000 was spent out of the district budget on kits for each grade level in each school, 20 in all, with an additional $1,000 from each school's budget. The parents raised additional money to buy scissors for each class, color-coded like the painting kits, according to the color for each class in the Art in Action curriculum. The other major cost was half-time release for Campeau from his high school teaching to act as consultant to the Art in Action program and to the assessment.

Of the five Bozeman elementary schools, only Longfellow, with its traditional orientation, wanted to evaluate student knowledge of art for the OTE program. They believed it was necessary to use some form of objective assessment to evaluate the success of Art in Action, as they would evaluate any OTE project. But it was not easy. There were no commercial tests and in any case the school could not afford to buy them or pay for administration. The assessment had to be a homegrown product.

The Longfellow OTE leadership team, comprised of Meagher, three Longfellow teachers, and some parents, sat down with Campeau in the late summer of 1989 to devise an assessment which would assess the program's success in teaching children to think about art as a construction, not as narrative, and to use the vocabulary of art appreciation.

They chose a pre-post design, with interview questions to be administered by parent volunteers. Each child was to be assessed, although interviewing a sample of children would be adequate for a program assessment. But young children do not like to be singled out and asked to leave the classroom unless they know everyone is going to get a turn. The designers of the assessment were also interested in looking at a full range of possible responses, since the vocabulary the children used was as interesting to them as the scores.

Section III. Experience.

Because of this history and context, the Longfellow assessment might be seen as a curious (and entirely unintended) parallel to a school's relationship with test publishers. The art assessment was designed because the Longfellow School principal felt the need for an evaluation of the program and could not find or afford a published test, so it replaces a test in the minds and attitudes of the school personnel. The teachers are not part of the process, since the OTE leadership group wanted to spare them additional work. The assessment was designed in consultation with the principal, a committee of three teachers, (and some parents), just as a test would be chosen by the principal and a few teachers. The parents fill the role of external scorers, (somewhat ironically), and the results were expected by the teachers with a tinge of the same anxiety prompted by test scores.

Some (although by no means all) of the problems experienced--unreliable volunteer help, resentment among parents asked to volunteer again when they thought they had done enough, resentment among teachers who didn't receive the results of the tests as soon as they would have liked them--are attributable to treating the situation as if it were the same as applying a published test, but without replacing the organization and expertise that support a test publisher's activities.

The assessment was designed to be conducted in the same way for both pre and post-test: a child is asked by a parent to look at a picture and "tell me everything you can about it." The parent uses only two prompts: "Are there any other things you can tell me about this work of art?" and "Is that everything you have to say about it?" The parents write the child's answers down on prepared sheets (see appendix), on numbered lines. The November tests were put away until May and then both tests were scored at a sitting by a group of parents.

It proved difficult to find pictures that would not stimulate children's desires to tell a story. The choice of pictures, neither too abstract to baffle the children nor too representational, remains a problem needing research and experimentation.

The test in school year 1989-1990
During the first year of the program, five parents did all the interviewing and the scoring, voluntarily, since there was no money for releasing teachers.

The pre-tests were given in late November 1989 and the post-tests in May, 1990. November was a little late, for instruction using Art in Action began in September with the return to school, but there had been no time to complete the form and train the parent interviewers in the push to get the program running in the schools.

The pictures chosen in the 1989-90 school year came from the Bozeman High School art collection and were all landscapes, which in the case of a seascape selected for the second grade caused an unforeseen problem: 90% of the children had never seen the ocean, and interpreted a curling wave as "snow coming over a mountain," when they could understand it at all. The pictures used for the interviews were different for the pre and post-test, which would be difficult to defend as a testing practice unless they were chosen carefully to have the same features and qualities. This is an example of the need for expert advice if the Longfellow art assessment were to be used rigorously.

The assessment in September 1990.
During our site visit, we witnessed parents interviewing children in all the four grades for the second year's pretest. The physical set-up was about the same: in a room separate from the classroom (usually the utility room where supplies are kept and materials prepared), the parent

prepared a comfortable corner. The picture was propped up on a chair, about eye-level to a child. The parent sat beside the picture, with the response sheets on her lap. A child sat in a small chair opposite the picture, answered the questions, and then left with a message to send the next child in. Since this was a pre-test, the interviews did not take long, and indeed all of class consisting of 25 students could be covered in about one and a half hours.

The testing was scheduled at this time in order to avoid the late start of 1989, and also to accommodate several schedules, including ours. As we have mentioned above, Campeau could not dispose of any resources at Longfellow School, so there was no liaison between him and the parents and teachers at Longfellow. This lack of a coordinator at Longfellow School led to misunderstandings about who was assessing a certain grade at a certain time, and similar confusions. Teachers apparently knew the date of the Fall 1990 testing only the week before it was scheduled. They had not yet been told the results of last year's testing, a fact they frequently mentioned and were clearly unhappy about.

Parents were again asked to administer the questions, a request which seemed to arouse minor resentment. Because of fears that parent volunteers would burn out as a result of so many demands on their time, Campeau had promised high school students as interviewers. However, no high school students appeared, perhaps because of the organizational difficulties mentioned above. One volunteer is also an aide in the fourth grade class in the morning, so interviewing meant that she had to stay all day at the school. She told us that the parents had been asked again "because you were coming," "you" being the visitors from Washington D.C.

The pictures chosen for this year's pre-test were all impressionist landscapes painted in acrylic by a local Bozeman artist, Freeman Butts. They were characterized by broad, swift strokes of color; diagonal placements of linear features like rivers or fences; contrasting textures, from multiple layers of paint to bare unpainted patches; and dramatic expressions of mood. Campeau had chosen them deliberately in order to focus students on the formal, technical, and expressive properties of the work, not "tell stories" about the scene as they had frequently done when faced with landscapes in the previous school year.

There were five canvases. One was designated for each grade, but when an extra parent volunteer showed up, a fifth picture was hastily found, to make two for first-grade testing. The picture was chosen hastily, again an example of lack of organization.

The Freeman Butts pictures were not easy for the children to understand. As we listened to their responses, it was clear that they were trying to relate the pictures to recognizable objects or scenes. One of the first-grade pictures was particularly hard for its intended viewers. It was smaller than the others, in an ornate frame, and depicted autumn trees seen across a diagonal yellow river. It did not offer a figure or object for a young child to connect with.

The record sheets were exactly the same as for the previous year, with one additional sentence. The parent said to each child: "I want you to look very carefully at this work of art and tell me everything you can about it. Think of how you talk about art in your art classes." The second sentence was added in order to focus the child's mind on art classes from the previous year. The expression "work of art" was used so that the child might supply "painting," "drawing," "watercolor," if possible.

As observers we may also have inadvertently made children shy and hesitant to talk, although the parent interviewers assured us that they have found a wide range of responses, including the silence

we often witnessed. Grade 1 children, even when faced with the less difficult of the two paintings, seemed unsure about what to say. One little girl listened to the parent's request and then asked : "You mean what do I think it is?" She finally said she thought it was pretty. Others identified hills, bushes, trees, green fields, snow. A number listed a few items of this kind and then said "That's all."

Grade 1 children are completely new to instruction in art, since Art in Action has no kindergarten curriculum. Their responses are baseline information. Children in all the other grades had experienced the curriculum in the grade before, even if they had transferred from another Bozeman elementary school. Some seemed to be trying to recall the terms they had been taught, but the three-month summer hiatus had made their memories hazy. Occasionally they would get flashes: "Oil on canvas, maybe?" said one fourth-grader. Another thought the brushstrokes were smooth and "somebody's been scrubbing some of the parts." Other fourth-graders were more effective at either remembering their previous instruction or responding to their landscape: "He mixed a lot of colors before he put them on. The paint's pretty thick. He used a thick brush." "It's splatter-painted, with lots of texture, mixed colors, thick, not small lines, brushed in long strokes."

On the whole, the children's responses were unremarkable, but their lack of distinction may be due to our presence, or to the kind of paintings chosen, and/or to the usual summer drop in retention of knowledge from the previous year's instruction. The parents remembered the children as talking more last year, but whether for the pre or post-test was not clear.

Section IV. Impacts on teachers and other personnel

Teachers

Talking with the teachers was complicated by the fact that the assessment impinges on them less than the art curriculum. If a question was not phrased carefully to refer to the assessment, it might be interpreted as referring to the curriculum.

They were all "sold on the curriculum," as one teacher put it. During a summer workshop for the district's elementary teachers, Campeau and the teachers adapted the curriculum to Bozeman's needs. He helped them to use a visual scanning chart and to understand that the material of art is not the object represented: the picture is not a horse but an arrangement of light, shadow, color, line, placement that causes the viewer to think of a horse.

Campeau firmly refused to support buying textbooks for each student: "then it would become like math," he said. He wanted the teachers to talk with the students about the sensory, technical, formal, and expressive qualities of visual art, using the language of art, instead of assigning them pages from a textbook.

The teachers found visual scanning a valuable technique in training the students to look at the art, not at the represented object. One teacher said that the curriculum had caused her to learn a lot by the end of the year: "It's not just artsy-crafty stuff," she said.

The teachers were not asked to participate in the assessment because Campeau, Meagher, and the OTE leadership team thought that working on the assessment would be too much of a burden for them. Their non-participation, therefore, was not exclusion but consideration for their extremely busy lives. "We said to them," Campeau told us, "if you concentrate on teaching this, we will do the evaluation--we will not make it another addition to the work that you have to do."

However, the teachers expressed anxiety about not having received the results of last year's assessment. (A parent had told us that teachers were apprehensive that they would be judged by their students' scores.) One said she didn't really know how the art curriculum was doing, because she didn't know how her students had scored--someone wasn't sharing that knowledge. She thought that the real strength of the assessment was in the use of the students' own language.

Another teacher also thought the teaching "did a world of good," but professed not to know how it was going since the scores had not been communicated. The teachers wanted to know what parts of the curriculum had not been successfully taught, so that they concentrate on those areas.

When we asked Campeau why he, Meagher, and the OTE leadership team had not told parents and teachers the test results, (although they had given them to us), Campeau responded that he wanted to avoid comparisons among teachers of the same grade (especially with the first-grade teacher whose students showed minimal growth), and he also wanted teachers not to teach what looked like weak areas, but to continue to teach the subject as a whole. He expressed some discomfort with having recorded the grades by individual classes, and wished they had aggregated them by grades, without having any record of the scores of individual teachers' classes.

In the end, we recently learned, he gave the score sheet (figure 1) to Meagher, who communicated them to teachers individually. He did not give them the sheet which appears in figure 1, but only the figures that applied to their class, so that they could see the amount of difference their teaching had made.

A polished assessment would involve teachers and would pay them
for any extra time or provide substitutes for released time. The Longfellow program is missing out on the feedback loop which is a vital feature of good performance assessments, when teachers are involved as designers and scorers of assessments. In another of our case-studies, the South Brunswick NJ K-2 Portfolio Assessment, teachers not only observe their students' literacy development, but also conduct interviews with the children and their parents.

However, Longfellow Elementary School is a traditional school, whose success on accepted measures of achievement is satisfactory. There is no evidence of interest in changes, especially not in the roles of teachers. The art assessment at Longfellow Elementary School, as we pointed out at the beginning of this study, was treated by everyone involved as much like a test purchased from a test publisher. None of the teachers expressed interest in its design or administration of the assessment, with the exception of saying that it did not disrupt classes.

While assessment in one first-grade class was proceeding, a teacher was using the Art in Action curriculum. She was teaching curved lines, holding the big Art in Action reproductions on her knee as she spoke to the group clustered on the floor around her. The children could see the curves in Hokusai's famous painting of the wave and Mount Fuji easily enough. They could also see them in Durer's woodcut of a rhinoceros. The teacher read the information on the back of the reproduction ("Durer never saw a rhinoceros, but drew the animal from notes he read"). She did not pick up on one child's comment that the rhinoceros's plated hide looked "like metal." She sent them back to their individual desks to make pictures with curves in them. Since the Art in Action reproductions are flat and without texture, the child's comment showed ability to appreciate texture.

Parents

Parents were involved in all aspects of the assessment. They worked with Campeau, Meagher, and

Fig. 1

PRE-POST TEST RESULTS FOR OTE GOAL FOR LONGFELLOW SCHOOL 1989-1990

Goal: To measure student knowledge in the art area and to show statistically significant improvement in both art knowledge and appreciation between pre and post testing.

Description of Test Protocol

The test was developed by the OTE leadership team composed of Lynn Hall, Nina DiMauro, Sharon Smith and Mike Meagher with the assistance of Ray Campeau.

Over the course of several meetings, we made the following decisions:

1. We would use parent volunteers to do the testing.
2. They would test every child rather than using a random sample.
3. In order to control the test environment and standardize it as much as possible, it was decided to ask each child only one question -- "I want you to look very carefully at this work of art and tell me everything you can about it."
4. We allowed two prompts -- Prompt #1 was "Are there any other things you can tell me about this work of art?" Prompt #2 "Is that everything you have to say about it?"
5. The pre tests were given in late November and the post tests in May.
6. Both tests were scored after the post tests were completed.
7. Both tests were scored by giving one point for each accurate comment that had to do with sensory or technical properties such as line, shape, color, medium, tools, etc. Two points were given for each comment to do with formal properties such as unity, theme, variation, repetition, etc. Three points were given for each expressive response, i.e. character of work, feeling of work, mood, etc.
8. The scores are derived from the number of points accumulated on each response page.

Test Results

First Grade	Pre Test	Post Test	Growth	# of Students
McCulloch	110	155	45	23
Reed	41	73	32	24
Dreiseszun	83	90	7	22
TOTAL	234	318	84	68

Second Grade	Pre Test	Post Test	Growth	# of Students
Hendricks	30	146	116	24
Manceau	83	150	67	24
TOTAL	113	296	183	48

Third Grade	Pre Test	Post Test	Growth	# of Students
Backer	109	161	52	23
Hall	80	125	45	25
Pierre	140	216	76	28
TOTAL	329	502	173	76
Fourth Grade				
Mecklenburg	36	126	90	24
DiMauro	22	179	157	24
Branting	28	162	134	26
TOTAL	86	467	381	74

My first reaction to the scores is a very positive one. I'll be very curious about the same testing for next year. All the classes made significant improvement with the possible exception of one of the first grade classes. I wonder why the growth is so much greater at 2nd, 3rd and 4th grades. I'll need to talk to the 1st grade teacher and to look carefully at the curriculum. Ray Campeau may also have some ideas on that one.

the OTE leadership team to design the assessment; they administered it; and they scored it. The parents involved in the design were especially helpful in the wording of the prompt (see appendix), because they knew they would have to use the words in interviewing the children.

Parents support the program, but, like the teachers, are inclined to see it as an instructional program and seemed not to judge the assessment separately. They were sensitive interviewers. One said that children should be close enough to touch the work of art, in order to appreciate the texture (one child did tap the painting, and then looked at his finger, as if he expected the paint to come off), and used gestures to invite them to touch. They criticized the original question as not specific enough, and the language as a bit too formal. One parent noticed that the younger children seemed unsure of themselves during the pretest, but relaxed in the post-test because they knew what was happening. "The older students liked the one-to-one attention," she added.

This parent was also eloquent about her reasons for enthusiasm: "In the past students were graded simply on their ability to draw. This new assessment centers around evaluating a student's knowledge of the properties of art (not everyone can draw--but <u>everyone</u> can enjoy art.) I like it."

Parents were trained for a session lasting about an hour and half before they scored the results in May 1990. Campeau explained how the scoring should go, using examples. Throughout the scoring, there were frequent pauses for discussion in order to maintain the same scoring standards.

For the post-interviewing, in May 1991, Meagher wants to replace parents with art education students from Montana State University. He believes the parent volunteers have burnt out, an impression which could be suppported from some of the remarks we heard in September 1990. Campeau wants to invite those parents who are interested to continue their association with the program, and will supplement their number with Montana State University students for the interviewing and for the scoring of both pre- and post-interviews.

<u>School district personnel</u>

The Bozeman School District Administration was enthusiastic about the art assessment and seemed to regard Campeau as a district treasure. We spoke with Paula Butterfield, Superintendent of Schools, and Brian Dunn, assistant superintendent. They were interested in other kinds of assessment than multiple-choice, but would find it hard to change in face of the highly successful record of the district's schools.

The chairman of the Bozeman School Board, James Screnar, was also supportive, although lacking specific knowledge of the assessment. Since the art assessment is school-specific and did not require line-item funding, no Board approval was required and no presentation was made to the Board about it. He was interested in assessing how well graduates of the Bozeman schools do in the outside world and supported Ms. Butterfield's move towards involving the business community in the schools.

<u>Section V. Scoring</u>

The scoring scheme is unsophisticated, but was consciously designed in that way by Campeau. He modeled the scoring activity itself on his own experience as a "reader" for the Advanced Placement Studio Art Portfolio Evaluation (see case-study IV): "They make me feel pretty good about doing this kind of thing." He did not want to get involved in control groups or elaborate reliability measures: "that isn't what I'm in to."

The scoring consists of points given for accurate comments, thus:

- one point for a comment on technical properties such as line, shape, color, medium (watercolor or oil), or tools (brush or knife, e.g.);
- two points for a comment on formal properties such as unity, theme, variation, repetition, etc.;
- three points for a comment on the expression of the work, such as its feeling, mood, character.

The points were marked at the end of each line, tallied for each student, and then for each teacher's class. The scores for pre and post were compared and a "growth" score computed.

The scores were charted according to each teacher's class (a matter of controversy, as we have mentioned). Growth ranged from 7 points of growth in one first-grade class to 157 in a fourth-grade, with an average of 75 points per class (fig. 1). Growth seemed markedly higher in the second, third, and fourth grades. The abnormally low score in one of the three first-grade classes (it was the only score lower than 32 for a class) was attributed to the administration of the assessment in a corridor instead of a quiet back room adjacent to the classroom.

The results of the pre and post-assessments seemed positive, but it was decided to do it again this year (and for subsequent years) in order to plot trends. While each child will accumulate knowledge over the years, only growth over a year is scored, so while the pre-test score may go up (because of continuing instruction in Art in Action), the distance between it and the post-test score must be maintained if the instructional program is to be judged a success.

The scoring of the assessment is unsophisticated, but adequate for the information required, that children had learned how to talk about works of art between the pre and post test. The differences in the growth scores may be attributed to the different developmental stages of children in four grades, which was noted by the principal in his notes to the results. Because 1989-90 was the first year the Art in Action program was in the school, the pre-test scores for all grades were low.

Since the scores are only indicators of growth, the evidence on which the scores are based--the recorded responses of the children--are of at least as much interest in showing the kind of growth that was observed. It is clear that the children learned how to use the vocabulary of art appreciation, in proportion to their grade level, for the later grades gained more than the first graders.

First Grade

The first graders began the year by describing the picture literally and commenting on its reality. "It looks real," "It's pretty," "I see green over here," "Mountains like we have, and water on mountains like we have," "There is a reflection on the water." The more talkative students usually named every color they could see in the picture.

Growth over the year was strong, and primarily due to many more children naming all the colors in the picture. "Little bit of purple and blue on the mountain...Black coming down stream. Maroon right there. Orange all over it. Little bit of yellow and green, dark pink and red in the pink." The literal descriptions became more detailed, although this is not what the Art in Action program

emphasizes. "Two squirrels. Crack in mountain. Tall cliff.... Fish, bird, and Indian there."

Of course there were children who said nothing during the pre-test and only one or two descriptive words during the post-test but even that shows growth. One child in both pre and post test became fixated on the ability of the artist "Art is always beautiful. Artist always draw beautiful."

Second Grade

By the spring, second grade children began to talk about techniques used by the artist. "Scratched in places," "Rock detailed because of bumpiness," "Mixes colors. Used fingers and a paintbrush," "The artist used a big brush, and fat brush, and a skinny brush." There was attempted discussion of shadow. "There is light and dark in the water," "It seems very dark." And some talk of perspective. "[They] sat on the shore and watched."

The main focus however, was still on the reality of the piece and the colors used. Because the work of art used in the post-test represented the ocean, which many of the children had never seen, there was some confusion. "I think it's mountains." Barring the confusion, the class on the whole seemed comfortable talking about the picture--what it was, what they thought it looked like, the colors represented. "Blue in water and white. Light blue in sky with orange. Purple. A little green. Some black." One student however, after an average response on the pre-test, said nothing in response to the picture of the ocean.

Third Grade

By the spring, third grade children were able to discuss technique. "It looks like they used thin and wide brush strokes." "To make the trees look like they go around its dark to make look shaded." "Right here it pokes out." "One side of the mountain is lighter than the other because of the shadows." "The paint on the trees and mountains kind of gloppy, kind of sticks out." "It looks like things in the front are painted really dark greens or blacks and the things in the back are painted orange red or yellow." "It looks three-dimensional." It is interesting to note that they are speaking in more complete sentences and are beginning to articulate clearly what they are seeing.

There were still a few students who tried to describe what was represented, but the number diminished.

Fourth Grade

The pre-tests of the fourth graders showed responses similar to the pre-tests of the first graders. Because 1989-90 was the first year they were exposed to the language, their age gave them no advantage. "It's a bunch of trees and fence. Bunch of different colors." "Just some flowers." "Looks like its fall." "Looks pretty." "Don't know."

There was a perceptibale gain in the post-test--the fourth graders were able to express their thoughts in complete sentences and to move beyond the literal image of the picture. There was also much more sophistication in the discussion of technique. "The hat look 3-D because of texture." "Suggests glasses with little bits of circle." "Used rough strokes, not little ones." "Looks like a wide brush was used." "Lip has light highlights so it shows up from the dark around it." "Some of the painting looks like dry brush, some like sponge." "Looks like sun shining it makes colors. One side of coat darker than other."

Throughout the grades, there was very little talk about mood, "Lots of warm colors" and "It looks mysterious," were the only phrases that referred to the mood of the piece. While the students knew that the artist is "painting to show how someone expresses his opinions," they ventured no guesses

as to what exactly the artist might be trying to express. In the spring, while some need to define the image still lingered, students usually worked past that and focused on the style (watercolor, oil, etc.) and technique used in the work of art.

The same amount and kinds of growth may be seen in the 1990-1991 school year, although the post-test is outside the scope of this study. Of great interest from the 1990-1991 scores will be the amount of growth recorded in students who will have spent two years in the Art in Action program by then--the second, third, and fourth graders. From our observations, the 1990 pre-tests showed little evidence of retaining vocabulary over the summer, but learning may have increased later in the year as students began to recognize the same terms and the same ideas at the next grade level of the Art in Action curriculum.

Section VI. Costs

The Longfellow Elementary Art Assessment Program costs more than appears on the surface. Ray Campeau was released from his position at the high school for a half-day for a full year, 1989-90, to set up the program and the assessment, so the cost of his time must include that of a half-time substitute during the school year. On the assessment itself (as opposed to the curriculum), he spent about four weeks of his 50% released time.

The parents' time spent in training for the assessment, interviewing the students, and scoring the responses was donated, but if the program were replicated using teachers or other non-volunteers, the costs of their time would probably be the largest part of the budget. The costs of the time spent by the principal, the three teachers, and the parents designing the assessment with Campeau would also have to be calculated to get a realistic sense of what the art assessment actually cost. The group met for about two hours on three occasions.

As it is, the assessment is regarded by all concerned as almost without cost, since the only financial transactions were transfers for Campeau's salary.

The money spent for art materials is not part of the assessment budget, since they were not used for the assessment.

Section VII. Impacts on minorities.

Since minorities are almost entirely lacking in Longfellow Elementary School, the question of impact did not arise. The only comment about possible differences between groups came from a parent, who noted that "poorer kids are more inhibited--art and talking about art is less part of their background." This was the only comment we heard in response to questions about possible differences in performance between races and groups.

Section VIII. Discussion

Longfellow School's assessment of learning about art suffers from the defects of pioneering programs put together on a shoestring budget. The assessment is not nor does it pretend to be psychometrically sound. There is no attempt to standardize details of the interviews, such as the child's distance from the painting, whether or not the child is invited to touch it, or the amount of time given to each child. No account is taken of other variables than classroom instruction: the

devlopmental stages through which children pass from first to fourth grade; the learning about art that children whose parents are artists or who spend a lot of time around artists may pick up of the school curriculum; and the differences among the expertise of the teachers.

However, it has some good points that might be made better by others taking up the idea and improving on it, as imitators frequently do. Its use of parents as interviewers is inspired, since it solves three potential problems at once: parents are involved and support the assessment; the students are comfortable because parents (frequently friends and neighbors in this small town) are not threatening; and teachers are freed from an additional assignment.

The project (and especially Campeau) has the courage to face important questions: do children in the lower grades have the cognitive skills necessary to enter into and understand works of art? Perhaps not on the more abstract level, although leaps of growth were recorded. (It will be interesting to see how the first graders from the first year discuss art when they reach the fourth grade.) What happens when children are taught the vocabulary and encouraged to use it? Does simply exposing children to the language of art when young enable them to grow into it when they are ready? These are questions that should be pursued. Simply to raise them is to stimulate thought about the learning of art.

Another important point about an elementary art learning assessment must be made: it has an important role in elementary education. Art is neglected in American schools as not essential when compared to literacy and numeracy, although there is a considerable literature affirming the value of art in developing both skills and judgment, not to mention cognitive and aesthetic faculties. Evaluating art routinely would mean that it would be taught and not merely given a few minutes of painting as a treat on Fridays. There is a chicken-and-egg relationship between the teaching and assessing of academic subjects in American education: what is assessed is what gets taught.

The major problems we observed were connected with the fact that the organization of the assessment was carried entirely by Campeau, who had other duties that affected his efficiency. The promised high school or college student interviewers did not appear, causing some resentment among parents who felt burdened by the demands of interviewing. A school contemplating an assessment of this kind--whether of art or any other curricular area--should appoint an on-site coordinator to keep teachers informed, schedule interviewers, and make sure that paintings and interview rooms are available.

Insufficient experimentation with the kind of painting suitable for the assessment is a problem which perhaps needs the attention of researchers who would have time to consider it fully. Campeau is probably right in not wishing to use representational works, but the choice of the Freeman Butts paintings seemed to stem from their availability rather than their suitability for young children. Time should be invested in trying out different works with these age groups to see which works of art focus children's attention on them as artefacts rather than as narratives.

The influence of a strong personality on a school program raises questions about the program's dependence on a single person. It is doubtful that the art assessment would survive Ray Campeau's departure, for whatever reason. The parents work well under his direction and have excellent ideas for revising the assessment, but none seemed to have the dedication required to continue it without his leadership.

In sum, a sound elementary art learning assessment might look a lot like the Longfellow program, but with attention to design and scoring based on psychometric research. It would identify a group

of works and put them in a sequence from kindergarten upwards, so that students would be led through a progression of learning how to understand the constructivist nature of art. The assessment would be standardized through uniform conditions of administering the test, and through a good deal more training for the interviewers. The assessment needs the cooperation of a testing expert sympathetic to its purposes and its relationship to the instructional program.

The groups now contemplating national or federal assessments would do well to encourage local responses to evaluation challenges. There is an abundance of psychometric skill at the national level which could be applied to creative ideas arising from local situations. Marrying the two might produce a sense of cooperation among levels of the U.S. educational system, now conspicuously lacking.

APPENDIX

SITE VISIT AND CONTACTS,

LONGFELLOW ELEMENTARY SCHOOL ART ASSESSMENT

Ray Campeau was host and guide for Kathleen Fulton and Ruth Mitchell for their site visit, 17 and 18 September 1990. During the visit, we observed as parents interviewed students in grades 1 through 4 at Longfellow Elementary School; interviewed the principal; and spoke with teachers involved in teaching the curriculum, as well as the parents, in order to record their thoughts and feelings about the art assessment. We also visited the superintendent and central office staff and spoke with a school board member. As with all site visits, we sent questionnaire ahead so that our interviewees would have a chance to think about our questions.

Ray Campeau
419 South Grand Ave.,
Bozeman MT 59715
406-587-4392

Paula Butterfield, Superintendent,
School District #7,
405 West Main St.
Bozeman MT 59715

Brian Dunn, Assistant Superintendent
same address

James Screnar, Chair, School Board
same address

Michael Meagher, Principal
Longfellow School
516 South Tracy
Bozeman MT 59715

Louis Gappmayer, Principal
Bozeman Senior High School
205 North 11th Ave.
Bozeman MT 59715

Shirley Talbott, Parent
Longfellow School
516 South Tracy
Bozeman MT 59715

Patti McCartney, Parent

same

Cheryl Kemp, Parent
same

Nina DiMauro, Sharon Smith, Lynn Hall, Donna McCulloch, teachers
same.

LONGFELLOW SCHOOL

Student #___ Name _____ Grade Level___Time_____

Question : I want you to look very carefully at this work of art and tell me everything you can about it. Think of how you talk about art in your art classes.

Prompt #1: Are there any other things that you could tell me about this work of art?

Prompt #2: Is that everything that you have to say about it?

Thank you for helping me.

Responses:

1._____

2._____

3._____

4._____

5._____

6._____

7._____

8._____

9._____

10._____

11._____

12._____

13._____

14._____

15._____

16._____

17._____

18._____

19._____

20._____

21._____

A Program with Distinction

ay Campeau's art classes at Bozeman Senior igh School are attracting national attention

igh School art instructor Ray eative suggestion to one of his

Ray Campeau walked from palette to palette last Wednesday in a forest of easels during "Painting and Sculpture," a class for second-year art students at Bozeman Senior High School.

As he strolled down a narrow aisle between both hurried and meticulous paintings, the high school's senior art teacher stopped and stood next to one of 34 students to critique a painting of three brightly colored figures spaced far apart.

He suggested she move the figures closer together and create a group, making the painting more whole. Campeau never told her what to do. He merely suggested.

The 15-year-old student agreed. She said later she was not dismayed and was still confident her painting would work.

This is the fourth painting the class has done after only nine classes. Campeau wants his students to churn out work and develop their own style. He asks students to "look for the essence" in their painting and to distort abstract images deliberately, not accidentally.

"I want them to get help from me but not get indoctrinated by me," Campeau said between critiques. "My energy is with the students and what they do with themselves. I could make all of these paintings look like mine but I try only to answer the questions that they ask."

Campeau, 57, got his first job teaching art in Bozeman schools. He has never left and has developed an art program that has garnered a national reputation.

Campeau, who begins his 31st year of teaching this fall, will be honored next week in Newsweek magazine's education issue. A Newsweek photographer took "about 500" pictures during a recent visit.

However, the nationally recognized art program only garnered a paragraph on page 41 of next week's issue. There are no pictures, according to a Newsweek spokesman.

Campeau and his students work in the ultimate studio. The school district converted its old works into an art studio several years ago. A sliding gara door serves as part of the studio wall, while shards stained glass rest in boxes marked "Quality Thin Sliced Meats."

The high ceilings, exposed brickwork and brigh murals on the walls offer a pleasant retreat from tl cinderblock hallways of Bozeman Senior High Schc which are also decorated with work from Campeau students.

Campeau changes the layout of the huge room every year to keep things fresh.

"The (school) district has respected this department," Campeau said. "It's not a dumping ground for kids. It's as hard to pass an arts course is to pass a physics course."

"He's incredible," Erin Frost, 16, said of Campeau. "He has just basically encouraged me. In high school, acceptance is such a big thing. The atmosphere (for acceptance) here is just great."

Frost, a teaching assistant in Campeau's second

BEST COPY AVAILABLE

Left, Corey Johnson steps back from his painting befuddled on how to progress. Johnson's work appears on the cover of 'This Week.' Below, Cagil Kurtman carefully draws in another character on paper.

year art class, has also delved into different mediums with Campeau's help, even when it wasn't working. "He'll make you see it through once you start it," she said.

Campeau admits that his students are not saints but knows that they are highly intelligent. "We try to make art on a very serious level," he said.

"I'd quit if I didn't like it," he said nonchalantly. "I'm pretty gregarious and I still get excited about other people's work."

The students can take courses in painting and sculpture, jewelry and print making and ceramics and stained glass. The terms is always split into two different mediums.

The program's success can be seen in its students. Bozeman is one of the only high schools in Montana to offer an advanced placement arts course. Here, students can take a college-level arts course for college credit in high school. The class will spend the year working on a portfolio, which will be judged by a nationwide panel of judges. Campeau will be one of them.

The highest score a student can receive on his unsigned portfolio is a five. In the past several year, Bozeman students have received four 4's and three 3's. "Fives do not come easy," Campeau told his class.

As Campeau walked around the room looking at the advanced students' work, a talented student with a fettish for sketching dragsters showed him his work. Instead of asking the student to change subjects, Campeau suggested he take tracing paper and sketch a picture of an oncoming dragster from another angle. "Make up a tool that looks like this," he said. "Try it from another angle and see how it feels."

"We're here to promote thinking," Campeau said. "I want a student to come out of here self-confident and understanding his own self worth."

Story by MAC DANIEL
Photography by DOUG LONEMAN
Of The Chronicle

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

VI. THE "WE THE PEOPLE..." CONGRESSIONAL HEARING CONTEST

A CASE STUDY OF A NATIONAL CIVICS COMPETITION

Section I. Introduction

The decentralization of American education encourages resistance to outside influences. There are many programs outside the education realm designed to link education to the real world, but the characteristics of the American educational system make it difficult for these outside programs to become part of the system. Among these programs are competitions such as the Odyssey of the Mind, the Physics Olympiad, the Westinghouse Science Talent Search, History Day, and--our focus for this case study--"We the People..., " administered by the Center for Civic Education, co-sponsored by the Commission on the Bicentennial of the United States Constitution, and funded by an act of Congress.

These competitions commonly have district, local, and national levels. The preparation for the competition may go on in the classroom (some programs, such as "We the People...," supply curriculum materials), and the competitions may be assisted by school district staff, but they are not officially part of the school program.

Many of these programs offer excellent examples of performance assessments, in contrast with the academic marathons and the "bees," (spelling, geography, and citizenship), which are tests of memory. For example, the Odyssey of the Mind asks groups of students to solve problems crossing academic disciplines. In last year's competition, sixth through twelfth graders were asked "to design and produce a transportation system consisting of five battery-powered vehicles carrying costume parts to each of the five, six, or seven teammates. Costume parts will arrive just in time to change the team's appearance. Vehicles will run in paths that will allow vehicles to crash into one another if their movements are timed incorrectly." Although they have months to prepare, the teams can spend no more than $75 on materials and equipment. Their final presentation at the competition is limited to no more than eight minutes.

The "We The People..." competition in particular, and other academic competitions in general, foster contact with the outside world. Students have the opportunity to see and meet people who now hold respected positions in the community. The link between schools and the outside world is probably the most neglected in the arena of school reform in terms of promoting a desire to learn

among students.

In the Center for Civic Education's "We the People..." competition, students simulate a Congressional hearing and respond to questions from a panel of community members. It has another important feature: "We the People..." is one of the few innovative assessments of history/social studies.

In the development of performance assessments, humanities have been slower to develop than those in writing and mathematics, the common subjects for accountability. In social studies, students read textbooks, answer questions, and take tests, mostly multiple choice. The tests only sometimes involve an essay synthesizing information from many different sources. Very little, if any, practical application of the learned knowledge is required at all.

We chose to observe the "We the People..." competition because it is one of the few high school level, history/social studies assessments. Its status outside the education system makes it an example that is often overlooked. We were also interested in the use of community volunteers as judges, a unique solution to finding "experts" to serve as judges.

Throughout the "We the People..." congressional hearing competition we looked at such issues as how the competition is experienced by the participants--teachers, students, and judges. What happens during the competition? How is the competition organized? Does the competition have potential as a means of evaluating the learning of history and government? Does "We The People..." have educational value? To whom is the feedback valuable--teachers, students, administrators--if anybody? Can this program model be used in other subjects--or interdisciplinary programs--can results be standardized and used in program evaluation? Answers to these questions will be found in the following discussions.

Section II. Purpose and Rationale of the Assessment

The purpose of the performance assessment in the "We the People..." competition is to judge the classes competing in the competition--to select the class which best uses what they have learned to respond to the questions posed by the judges.

The "We The People..." curriculum and congressional hearing competition was developed by CCE to enhance understanding of and interest in the democratic principles upon which the United States is founded. Because these principles are embodied in the Constitution and the debates that surrounded its inception, the content of the "We The People..." curriculum and competition is constitutional principles and their application to both historical and contemporary issues.

Many different forms were reviewed before the CCE decided on the congressional hearing as a form for the sorting mechanism. The primary consideration of CCE was that the program, both the competitive and non-competitive portions, foster co-operative learning. Co-operative learning is not simply groups of students working together to reach a goal. In order to maximize co-operative learning, the groups within a class must be inter-dependent. That is each group's success or failure must be dependent on all the other groups' success or failure. Another issue for CCE was that all other history/social studies competitions (for example, Mock Trial) involve a select group of students. CCE wanted students of many different levels of accomplishment to compete.

Essay tests were immediately ruled out as they did not meet the co-operative learning requirement. Then the competition was designed as a debate rather than a congressional hearing model. But because the hearing model allowed for a prepared statement and free response, CCE believed it

would assess two different competencies rather than one. CCE wanted to focus on the students' acquisition of knowledge rather than their presentation of that knowledge and felt the hearing design would allow for this better than a debate format.

CCE personnel decided to develop a system in which an entire class participated. Breaking the class into small groups ensured that each student would participate. It also meant that each group would rely on and watch the performance of the other groups within the class. In order to ensure equity, teachers must certify that their class was not constructed specially for the competition, although by allowing Advanced Placement classes to participate this rule can be bent, since AP classes are specially constructed of students regarded as capable of college-level work.

The "We the People..." competition focuses on group assessment as opposed to individual assessment. Group assessment is closely linked with program assessment--how effective the program is at conveying the content of the curriculum. As a program assessment, "We the People..." functions in an exemplary fashion. The assessment directly assesses the "We the People..." curriculum, providing accountability; the class is judged as a whole but on the participation of all; and the assessment models the type of instruction desirable, focusing on the synthesis and application of information. The only function it does not serve at the present time is individual student selection and sorting (certification). However, if the model were to be adopted as an educational assessment within the system, providing individual student scores would not be out of the question.

Section III.The Center for Civic Education's Experience to Date with "We The People..."

The National Curriculum and Competition

The Center for Civic Education (CCE) is a non-profit group which researches, develops, and implements programs of civic education for public and private schools, elementary and secondary.

The "We The People..." congressional hearing competition is designed so that high school students work in cooperative groups. All students participate in presenting the "testimony." The judges act as congressional committee members, who first hear the students' oral testimony and then ask follow-up questions to probe their depth of understanding and their ability to apply their knowledge. The judges collectively rate the student group according to six criteria: understanding; constitutional application; reasoning; supporting evidence; responsiveness; and participation. Following the competition, the scores are reported to the classroom teacher as feedback for the students and improvement of learning and performance.

The overall experience of CCE with "We the People..." has been encouraging. Those teachers who participate usually continue with the program. Most teachers and judges believe that this is an extremely effective way to teach the Constitution. Student testimony is glowing: some students rank the competition as the single most exciting event at high school.

The "We The People..." curriculum, developed by CCE in consultation with leading scholars and educators, complements the regular school history/government curriculum and is designed for a wide range of achievement levels. Curriculum for the "We The People..." program is available at three instructional levels: Level I (upper elementary), Level II (middle school), and Level III (high school). At each level, a variety of suggested teaching strategies are employed, such as class debates and individual research into important contemporary issues. This case study involves the Level III instructional material which is found the "We The People..." textbook. It is written in clear, simple language in order that many levels of achievement within high school will be able to use the

materials effectively. In fact, one student of an Advanced Placement United States History class said, "We thought [the book] used very simple language but we still found it very useful."

The "We The People..." text book was written by the staff and consultants of the Center for Civic Education with the assistance of scholars from the fields of political science, history, and constitutional law and teachers and curriculum specialists from schools throughout the nation. The text was published after extensive field testing and has received praise from students, teachers, scholars, and those in government for its accuracy, interest, and effectiveness in dealing with the history of the development of the Constitution and Bill of Rights and historical and contemporary issues surrounding these documents.

A sample unit is organized as follows. It begins with the purpose of the lesson and a list of ideas/words the students should know. The body of the lesson follows. The students are then asked to apply what they have read to detailed and complex questions. Occasionally there is a section detailing how the Founding Fathers would have responded to the philosophical questions posed to the students. Review questions end each lesson, reinforcing the important issues dealt with in the unit.

The simulated congressional hearings are an elective to the curriculum and can be either competitive or non-competitive. Competitive hearings require State and District Coordinators and precise adherence to the format designed by CCE. In addition, schools sometimes set up their own informal and non-competitive hearings, although in this case study we have focused on the formalized competitive congressional hearings.

The judges for the competition are community representatives, such as professors from local colleges and universities, attorneys, civic group leaders, and political office holders. "We The People..." also boasts a well-organized administrative support system ready to be drawn upon by those schools who wish to participate. The program is administered nationwide in 435 congressional districts, four trust territories, and the District of Columbia.

In 1990, CCE estimates program participation at 33,000 teachers and 2.2 million students. The overwhelming majority of the students (perhaps as many as 90%) did <u>not</u> enter a competition.

Over the four years the competition has run, there have been some adjustments to the format. Even before "We the People..." was introduced the program was pilot tested extensively. During the first two years the program ran, students were allowed to repeat: students who had participated one year were allowed to participate the following year. Understandably there was loud opposition to this, so CCE changed the rules to exclude any student who had previously participated.

The only disadvantage of the "We the People..." competition so far is the fundraising required to send the students to the state and national competitions. While the state and district coordinators do a majority of the fundraising, teachers (and students) must fill in the gaps with their own efforts. Elaine Craig of CCE says that this obligation keeps some teachers from joining the competitive portion of the program. "They are not fundraisers. They became teachers <u>because</u> they are not fundraisers."

<u>The Birmingham AL Competition</u>

The Alabama sixth congressional district "We The People..." congressional hearing competition was held at the Cumberland School of Law on the campus of Samford University just outside of Birmingham. We were guided in our visit by Dr. David Sink and Janice Loomis.

Dr. Sink is the Alabama State Coordinator for the "We The People..." competition and a political science professor at University of Alabama at Birmingham (UAB). He is also a civil rights historian and was able to give us an overview of Birmingham's historical and educational background. Janice Loomis, the Congressional District Coordinator and Director of the Alabama Center for Law and Civic Education, also helped to guide our visit so that we were able to observe the most activity in the least amount of time. It is the job of the State Coordinator and the District Coordinator to recruit schools and teachers to participate in the "We The People..." program and recruit judges to assess the "We The People..." competition. The State Coordinator has the added responsibility of serving as the liaison between CCE and the individual programs to ensure the program is implemented in the way CCE intended.

The State Coordinator annually attends two national conferences sponsored by CCE for the purpose of getting all the involved states together for training workshops and discussions of the "We The People..." program. The District Coordinator attends one of the conferences. Primarily the conferences provide training for coordinators in the content and implementation of "We the People..." programs and serve as a forum for state and district coordinators to share experiences and work out the kinks in their individual programs. The State Coordinators also meet with their members of Congress to encourage involvement in the program and maintain it in the public eye. CCE pays for the travel and expenses of the participants.

The day of the competition in Birmingham began with a judges' meeting to familiarize the judges with their responsibilities during the competition. The judges were community members from many walks of life concerned with constitutional law. Five of the Birmingham judges were attorneys from local law firms; others were the Assistant Dean of Continuing Education at the University of Alabama at Birmingham (UAB); three graduate students in Public Administration at UAB; a Professor of Political Science at Samford College; two Professors of Law from the Cumberland School of Law; a Professor of Education (Curriculum and Development) at UAB; the Assistant to the Vice President for Public Affairs at Alabama Power; an independent filmmaker currently producing and directing a documentary on the 25th anniversary of the Voting Rights Act; a Police Officer; a private consultant to local government; and an aid to Congressman Ben Erdreich. The judges for the district competition were recruited by Ms. Loomis. The activities of Dr. Sink and Ms. Loomis within the community have put them in contact with people concerned with both constitutional issues and education.

As the judges arrived they met with their teammates and discussed which of the three questions proposed for their unit they would ask. The judges are assigned to a unit and remain a team throughout the competition. Immediately before the competition, the unit judges decide which of the three questions to ask the students. Each judging team asks the same question for each unit in order to maintain some standardization. There are eighteen judges, three judges per unit. Groups of students appear before three judges at a time to respond to questions from a specific unit.

During the initial meeting, the judges were shown an instructional video and received last minute clarifications. The program then moved to the moot court where the students, teachers, and judges listened to Parham Williams, Dean of the Cumberland School of Law, welcome the "We The People..." competition. He reaffirmed the students' hard work and said that "win or lose, there is always a place for what you have learned."

The Birmingham sixth congressional district "We The People..." competition took two hours. Six time slots of 20 minutes allowed the judges to move from room to room, administer the question,

ask follow-up questions, and provide feedback to the students.

Because it is easier to have the judges move around rather than the students, each class of approximately 30 students remains in one room for the entire competition. The class is divided into six groups, one group specializing in each unit of the curriculum. The class competes as a team with each group taking its turn at the front of the room to answer questions. For each group of four or five students, the class is judged on six criteria: understanding, Constitutional application, reasoning, supporting evidence, responsiveness, and participation.

When the judges enter the room they do their best to create the atmosphere of a congressional hearing. They introduce themselves as Congressman or Congresswoman so-and-so and call to order the congressional hearing on the Constitution. The students are then told which of the three questions they will be answering. Their prepared response must be under four minutes. After the prepared response, the students must answer six minutes of follow-up questions from the judges to gauge their ability to think on their feet and apply what they have learned to current issues facing the nation. While they may use notes to help with the prepared response, they may use no aids during the follow-up questions.

The students were nervous. Their speech halted, their voices cracked. One girl was in tears a few moments before the competition began because she had dropped her notes in the restroom and they were wet. "Uhms," "y'knows," and "likes" filled dead air. While their ineptness as speakers reflects the pervasive lack of training in oral discourse in schools today, their responses were frequently intelligent and well-thought out.

We also noticed the excitement mixed with the nervousness, an excitement comparable to that found off-stage right before a performance or on the sidelines of a playing field before the game begins. The bright eyes and nervous chatter of the students was refreshing in that it was being directed at something academic rather than extra-curricular.

The questions asked were scholarly and required serious analysis in order to fully respond to the issues addressed. It struck us that few citizens of the United States, let alone high school students, could begin to intelligently and coherently answer the questions posed at the competition.

The schools provided us with valuable insight into the forces and issues affecting schools in the Birmingham area, although we are using them as general examples of schools who enter the competition.

Berry High School is a suburban Birmingham school which has participate in the "We The People..." congressional hearing competition all four years the competition has been held, usually finishing in the bottom half of the standings. This year however, the teacher taught the "We The People..." curriculum to three classes and held a runoff competition between them to see which class would enter the competition. The strategy paid off, because this year the Berry team placed second. No minority students were represented on this team.

John Carroll High School is a Catholic high school in Birmingham. All the students wore white shirts and navy blue pants or skirts. In 1988, the first year of the "We The People..." competition, John Carroll won the district division. Although the teacher has taught the curriculum each year since then, she has allowed her students to vote on whether or not they want to participate in the competition. Her 1989 class voted not to compete, so the 1990 competition was the first one John Carroll has been in since 1988. Two African American students were in the class. The class placed

third in this year's competition.

Corner High School was the only rural high school represented. Both the teacher and the students betrayed their nervousness about the competition in their mannerisms and speech. In fact, the teacher called Dr. Sink the day before the competition to tell him how scared both she and her class were and that they might not come. Dr. Sink managed to calm her down enough to convince her that after putting in as much effort as they had into the curriculum it would be a shame and a waste if they did not compete. The class included no minority students. The class came in fourth in the competition.

Shades Valley Resource Learning Center (RLC) is a magnet school which draws the most talented students from Birmingham and its surrounding areas. The school schedule is designed to allow students one day a week (Wednesday) to pursue an internship with a local business, with academics on the other four days of the week. Shades Valley students are involved in many different competitions and are constantly interacting with the world outside school. Shades Valley had the largest number of minority students, two African Americans and one Asian Indian. The class came in first in the competition.

While the schools competing may not have been indicative of the racial breakdown of the city of Birmingham (55% African American and 45% Caucasian), attempts were made to balance out other areas of the competition. Two of the competition judges were black, as were the three assistants who added up the final scores.

For the competition, classes of approximately 30 students divide into six groups of four or five students, each group making one of the six presentations and responding to the related questions. In preparing for the competition, the entire class must work together as a team, so that all students participate rather than a select few.

Ahead of the competition, CCE sends the participating schools the pool of possible competition questions with the instructional materials that accompany the texts. For each of the six units, CCE develops three questions. While the entire class studies each unit, one of the six groups takes responsibility for each of the six units in the "We the People..." text. Students prepare answers to all three of the questions for their unit but will be asked only one question at the hearing. While the students receive the primary questions in advance, they have no idea what the follow-up questions will be. Entire classes work cooperatively to prepare and present the statements and answer possible questions.

Section IV. Impacts on Teachers and Instruction

The curriculum is designed to be interwoven with the regular United States history curriculum and by the testimony of many teachers simply enhances rather than displaces the regular curriculum. Because the "We the People..." assessment scores students on their ability to apply information orally, some classroom activities are altered to provide practice for the students. The non-binding nature of the curriculum allows teachers to take what they want from the curriculum and leave the rest. Of course, what they decide to use will reflect on their students' performance in the competition.

United States history and government teachers are recruited by the "We the People..." State and District Coordinators to participate in summer workshops on the "We The People..." curriculum and

competition. Teachers who have taught the "We The People..." curriculum lead the workshops and students who have participated in the "We The People..." competition demonstrate the competition for the new teachers. Some advanced academic inquiry into the Constitution is provided to the teachers at the workshop.

Usually only a few of the teachers at the workshop choose to use the "We The People..." curriculum during the academic year. Although the six-week curriculum is designed to be interwoven with the regular history curriculum, it suffers from a mistaken perception that it is primarily for advanced classes and that it will take time away from other methods of instruction. Teachers who have used the "We The People..." curriculum think it provides greater understanding of why and how the United States government was created. One teacher noted, "In many ways it brings history alive for these kids. They begin to identify with historical figures as people. It's wonderful."

The summer workshop serves as teacher development for the State of Alabama (our focus for this case study). This is also the case for the majority of the schools participating. While teachers do not get paid to attend the workshops, their attendance does count towards their professional development credits. In Alabama, teachers must acquire a certain number of professional development credits per year in order to maintain their pay scale. In this way participation is linked to direct rewards for teachers.

The high school version of the "We The People..." curriculum is usually incorporated into the history or government curriculum at either the 11th or 12th grade. Because the entire class participates in the competition, the entire class has research assignments and must complete assignments relating to the curriculum. We were told that many students exceeded requirements, spending literally hours at law libraries reading up on law cases pertinent to contemporary constitutional issues.

Section V. Scoring Issues

CCE has developed a rudimentary scoring system to accompany the "We the People..." competition (See figure 1). The teams are scored comparatively. Because the focus of the judging is on selecting teams for the next round of competition, precision scoring is not as necessary as it otherwise might be. The number scored makes no difference in itself. It is how that score compares to the other teams' scores that matters.

Later in this section we discuss the possibilities of using the competition to assess the quality of the program or to assign individual student scores. If that becomes the goal of the assessment, then a more sophisticated system of scoring would need to be developed.

The judges are provided with the "We The People..." textbook several weeks in advance of the competition. They are expected to read at least the unit of the book they will be judging (and are encouraged to read the entire book) in order to understand what the students have learned and, on a basic level, how sophisticated their answers should be.

The morning of the competition the judges meet with the State and District Coordinators to discuss what is expected of them throughout the competition. They are shown a video which describes what they will be looking for in student responses and presentation. The video is produced by CCE and narrated by Charles N. Quigley, Executive Director of CCE. He begins by discussing the need for quality and fairness in the judging, reiterates CCE's goal to make the "We The People..." a valuable educational experience for the students, and stresses the judges' obligation to allow students to make their points. If the students finish their prepared statement in

Figure 1

National Bicentennial Competition on the Constitution and Bill of Rights

directed by the Center for Civic Education
cosponsored by the Commission on the Bicentennial of the United States Constitution
funded by an Act of Congress

Unit # _____
Teacher _____
School _____

Congressional Hearing Group Score Sheet

For each criterion listed, score the group on a scale of 1 to 10, with 10 being the best score. (Use a separate form for each group. A summary form will be provided for totaling each class's score.)

1-2 = Poor 3-4 = Fair 5-6 = Average 7-8 = Above Average 9-10 = Excellent

	SCORE	NOTES
1. **UNDERSTANDING:** To what extent did participants demonstrate a clear understanding of the basic issues involved in the questions?		
2. **CONSTITUTIONAL APPLICATION:** To what extent did participants appropriately apply knowledge of constitutional history and principles?		
3. **REASONING:** To what extent did participants support positions with sound reasoning?		
4. **SUPPORTING EVIDENCE:** To what extent did participants support positions with historical or contemporary evidence, examples, and/or illustrations?		
5. **RESPONSIVENESS:** To what extent did participants' answers address the questions asked?		
6. **PARTICIPATION:** To what extent did most group members contribute to the group's presentation?		
GROUP TOTAL		

TIE BREAKER* _____

JUDGE:_____ DATE:_____

CONGRESSIONAL DISTRICT:_____

* Please award up to 100 points for this group's overall performance. (Bonus points will only be used in the event of a tie.) Bonus points should be awarded using the following scale:

Excellent	90 to 100 points
Very Good	80 to 89 points
Average	70 to 79 points
Below Average	0 to 69 points

less than four minutes, the remaining time is added to the six minutes allotted for the follow-up questions. The judges are also told that they can add time on to the ten minutes if a judge asks an extremely long or complicated follow-up question.

The video concludes with instructions to the judges on what not to consider when scoring the students. Judges are to judge student teams on the content of their presentation only. No points are to be taken off for using notes during the four minute prepared presentation. The students have been instructed that regular school clothes are appropriate for the competition and judges may not mark teams down on the basis of appearance.

Differing opinions among the students on the teams are allowed, even encouraged. Judges should score on the basis of how well students support their positions and how substantial their response. Students are not to be penalized for an accent, dialect, or poor grammar. As long as students display an understanding of the concepts they are to be judged on that basis. Students may sit or stand depending on what is most comfortable to them. Feedback should be constructive, and judges are asked not to be cruel, but to always end feedback with a positive statement.

Each group within a class is evaluated as they respond to their unit questions and the scores of the six groups in each class are added together to determine the final score of the class.

The students are judged on a scale of 1-10 (1-2 = poor, 3-4 = fair, 5-6 = average, 7-8 = above average, 9-10 = excellent) on each of six criteria:

1. Understanding: To what extent did participants demonstrate a clear understanding of the basic issues involved in the questions?

2. Constitutional Application: To what extent did participants appropriately apply knowledge of constitutional history and principles?

3. Reasoning: To what extent did participants support positions with sound reasoning?

4. Supporting Evidence: To what extent did participants support positions with historical or contemporary evidence, examples, and/or illustrations?

5. Responsiveness: To what extent did participants' answers address the questions asked?

6. Participation: To what extent did <u>most</u> group members contribute to the group's presentation?

After the first year, on the advice of the judges, CCE changed the score sheet in two ways. Originally the scoring scale was 1 to 5 but the judges said they needed to be able to differentiate teams more effectively and so the scale was changed to 1-10. The other change was that "responsiveness" was added to the scoring criteria. The judges found that often students would give reasoned and articulate answers to follow-up questions that did not address the question asked, so they wanted to be able to reflect that in their scoring.

The judges are allowed to keep their score sheets until the end of the competition in order to change scores in comparison to other teams they see later. Before the score sheets are handed in each judge must award a score of up to 100 points for the group's overall performance (0-69 = below average, 70-79 = average, 80-89 = very good, 90-100 = excellent). This performance score is

only to be used as a tie breaker in the event two teams have the same raw score.

The scores are added up for each judge and each team. There is no attempt at averaging, so final scores can be quite high. Out of a possible total of 1080, the national winners of the 1990 competition scored 999 in the final round of competition. The scores of the Birmingham competition were in the 700 to 800 range.

Teachers may request a breakdown of scores, but the individual judges' scores are not released in order to eliminate scoring disputes while still providing feedback.

The scoring system is simple and transparent. It would need sophisticated weighting if it were to become a formal assessment. However, it would not be hard to get individual scores from the team's work: students could write individual essays based on the questions in both the prepared and extemporaneous portions. Trained teachers could use a system similar to that of the Advanced Placement program using standard setting sessions and norming sessions to ensure standardized evaluation of each student's paper.

Students' group interaction could also be judged by checklists filled out by both teachers and students, as is proposed in the Connecticut high school mathematics and science assessments, which are also designed for groups.

The competition is already roughly standardized, since the same questions are asked in order to make fair comparisons. It would require some technical scrutiny to make the scoring reliable, but since other scoring techniques such as the Advanced Placement Studio Art portfolio assessment have been developed with the assistance of psychometricians, it would not be impossible to produce an acceptable system.

Section VI. Economic Issues

To adequately dissect the economic impacts of the "We the People..." performance assessment would require a great deal of time and calucluation. We can only indicate what should be inlcuded in a complete analysis of its economic impact and attempt to provide a road map for future research.

In commemoration of the 200th anniversary of Congress, CCE introduced "We The People... Congress and the Constitution," a non-competitive companion program of the National Bicentennial Competition of the Constitution and the Bill of Rights. The National Bicentennial Competition of the Constitution and the Bill of Rights is funded by an Act of Congress and co-sponsored by the Commission on the Bicentennial of the United States Constitution. CCE competed with many organizations for the grant, which totals approximately $3 million per year over the five years, 1988-92, that the program is scheduled to run.

The bulk of the program's budget is allocated to:

- Support for state and Congressional district program administration;
- Training for social studies curriculum leaders and teachers; and
- Distribution of free sets of curricular materials to every Congressional district.

A one-time, major cost was the development of the curricular materials--the cost of researching and writing the "We the People..." curriculum.

The entire cost of the program is born by CCE and private contributions. Individual schools do not have to pay to participate. CCE is faced now with finding ways to continue funding the program after 1992.

Some support is given by CCE to teams attending the national competition, although the state program is primarily responsible for those funds through fundraising.

The judges volunteer their time to the competition: four to five hours the day of the competition and whatever number of hours it takes them to read the unit of the textbook they have been assigned and become knowledgeable about the level at which the students should be participating. Janice Loomis estimates this time at approximately three or four hours.

Volunteer time donated to the program does much to ensure its success. Considering that many judges are lawyers, university professors, and law enforcement officials, the economic benefit of donated time is apparent.

Section VII. Impacts on Minorities

CCE was at first apprehensive about using a competitive model for the assessment of the "We the People..." competition. They worriied about the negative effects of encouraging competition among schools and students. However, because the competition was designed as a cooperative venture Elaine Craig says she is "sold on the competition" and that the competitive format has "produced nothing but benefits." She believes that there is social risk attached to the competition which motivates students to invest efforts far beyond what they put in to their other school work.

The competition also provides schools with publicity for their academic achievement as opposed to their athletics and extracurricular activities. A school in New Jersey, after winning the state competition, was greeted when they returned home by the school marching band and all the ballyhoo associated with an athletic victory. CCE believes that it is important to reinstate the preeminence of academics in schools and that the "We the People..." competition is one step in that direction.

In the Birmingham competition we were acutely aware of how few minority children were represented; four African Americans and one Asian indian out of approximately 120 students. Dr. Sink attempts every year to recruit more minority schools for participation. He finds it difficult not because of a lack of desire but because of a lack of resources, "Teachers in those schools are fighting a war. It doesn't surprise me that they don't have time for this competition."

Some primarily black, inner city schools use the curriculum but do not come to the competition. Last year, one such school in Alabama was prepared to come and compete when, the day before the competition, race riots broke out in school. As a result, the principal confined all students to their homerooms and refused to let the history class leave the campus to compete.

During the Birmingham competition, students were called upon to provide examples of certain principles--such as the right to revolution. They chose their own examples, and their choices were revealing. Despite Birmingham's place in the civil rights movement, none of the classes chose to use

as an example of a citizen's right to revolution the civil disobedience used so effectively during the civil rights movement. None of the classes chose the civil rights movement as an example of a minority's successful attempt to change the will of the majority. At one embarrassing moment when a student tried to suggest that there were organizations who could speak for groups, he could not remember what the initials NAACP stood for.

In our brief visit, Dr. Sink showed us the park in which many of the protests took place, the trees still bare in patches from the fire hoses used on black protesters. We drove by the black Baptist church where the four young girls were killed when a bomb was thrown from a passing car. We drove by the Birmingham jail and, although it was not the one where Martin Luther King served his sentence, (Dr. Sink believes that jail has been demolished), it was no less a powerful image.

While students have the latitude to choose examples, the Birmingham students' lack of exposure to the recent history of their community somewhat narrowed their choices. The reasons for this are many and varied but we feel it is safe to say that the racial balance of the classroom may, in its own way, shape what is discussed and made available to the students.

In response to a follow-up question asking how to increase voter turnout, one member of the group responded, "more education about the government and how it works would help people understand how deep their responsibility is. I know that this class has taught me more than I ever knew before and that without it I probably would not want to be involved in politics either."

Section VIII. Discussion

The two most impressive qualities of the "We The People..." competition were the level of commitment of the students and of the judges. The students did it because they enjoyed it and to prove to themselves that they could. The judges volunteered because it was a way to give back to the community that which they had gained from it.

The combination of the public school system with the community could be a powerful tool for reforming and restructuring our nation's educational system. One of the criticisms of schools is education's lack of ability to make education real for the average student. There will always be students who strive to succeed academically for reasons having little to do with school--pride, emphasis on education at home, or a desire to go on to university or college. However, for the majority of students (even those who do succeed) high school is like doing time until they can get out and get a job or go on to college. With the support of the community, students are able to see what a sustained effort in education can accomplish.

Because the curriculum inspires such motivation on the part of the students, the curriculum may be appropriate for those whose learning is not motivated by other factors such as emphasis on education at home, desire to be admitted to a good college or university, or love of learning. Some experimentation was done with at-risk groups for the first time in the fall of 1990 although the small amount of data is not conclusive yet. Successful hearings have been run with classes from juvenile justice settings in San Diego, California: classes from El Cajon Summit High School, a community-based residential home for juvenile offenders; Phoenix High School, a branch of San Diego County Juvenile Hall; and Youth Day Center School Program, a school which provides an eight-week transition program before release into the community, all participated.

In each of the teams we viewed, the students took the competition seriously. In the weeks before

the competition they met outside of school and researched answers. They seemed to find it exciting and challenging, something towards which they could work with a clear purpose.

Students trained in the "We the People..." curriculum are able to speak about the government on a fairly sophisticated level. As observers, we were impressed by how much they knew and how willing they were to discuss and analyze information. The content and quality of their answers were high. Not scoring the students on their delivery was a deliberate act on the part of CCE, but their delivery needed development. Our observations speak to the lack of rhetorical training found in our nation's schools.

Teachers assert that the knowledge learned from the curriculum and the competition is not forgotten and is drawn on all year to supplement and enhance discussions about historical events. One teacher responded when asked if her students quickly forget the material once the competition is over, "Oh no, it becomes a background for the [AP US History] class. Over and over they refer back to such concepts as civic virtue or right to revolution in order to explain and put in context certain historical and modern events."

Because the competition does not originate with the school, administrators do not always understand what it is about. We found that if a teacher can make a good case for the curriculum, which most can, then they are able to teach it and to go to the competition if they think it will benefit their students. For those schools that win any level of the competition, there is the benefit of the honor, an honor which administrators are likely to appreciate.

Judges like participating in the competition because it allows them to keep in touch with what is going on in education. One lawyer said, "There is something satisfying in seeing all these students getting a charge out of the Constitution. It makes me realize that all the glooomand doom I hear about education is not necessarily the whole truth."

The competition appears to have enormous potential (as yet unrealized) as a model for the evaluation of history/social studies and government classes. It is the most imaginative and well organized social studies assessment we know of, more impressive than some ideas at the state level. The mystery is why it has not been seized as a model.

The designers of "We the People..." put together an evaluation which is faithful to the intent of social studies and government--understanding how our constitution works, its origins, and its application to present problems. It asks students to prepare a statement, thus evaluating their ability to present a case given the benefit of research and reflection, and then requires them to apply what they learned. This design seems exemplary.

"We the People..." is almost a type case of performance assessment. It is a direct demonstration of students' ability to do what we expect after studying the Constitution and the Bill of Rights--explain their history and background and apply them to other times and situations. The requirements for working in teams and the involvement of the community are not found in all examples of performance assessment, but this case underlies their potential value.

Why does this competition succeed in motivating students to work independently, take responsibility for their work, and constantly improve their performance? These are possible answers:

- There is some outside validation of students' ability and success.

- Responsibility for decisions about learning rest on students.
- Teachers act as guides rather than the definitive authority.
- The subject matters to the students because it has "real-life" status.
- The program is socially real. For example, performing in front of distinguished members of the community may be thrilling but it can also be terrifying. The fear of failure and desire for success are magnified, and the student is asked to take a social risk as well as an academic one. But students rise to the occasion. The attempt strengthens them and provides them with an experience to fall back on. "That was the worst thing I've ever had to do, and the best," said one student.
- The program's teacher training helps teachers instruct all children, not just the brightest and the most motivated.

As we pointed out earlier, the funding for "We the People..." will end in 1992. Rather than letting it expire, or continue in an attenuated form, perhaps a state testing director might choose to use it to assess history/social studies, and work with the CCE staff to adapt it for a statewide assessment. In its present form, with the requirement for whole class involvement and presentation by teams, it works well as program assessment, which is usually the scope of statewide assessment. But as we have said, it would not be hard to derive individual student scores with some additional information, which would make the competition useful also as district-level assessment.

Drawing the program and the competition into the educational system would ensure the formation of a feedback loop between assessment and curriculum. Because "We the People..." is voluntary now, only a few teachers and classrooms get the benefit of seeing clearly what has been learned and what has not.

To judge by these four Birmingham schools, learning in history/social studies needs the information that this assessment could provide. Universally, the students did better on the prepared than on the impromptu answers. Some of these were embarrassingly poor. It is possible that the combination of dumb downed text books, multiple-choice tests, and rote learning may contribute to the student's discomfort with extemporaneous speech. This would need further research.

"We the People..." is obviously a stimulating and positive experience for all concerned, including community members who give it their time for at least the equivalent of a working day. In Birmingham, the success of the program is alloyed with regret that it does not reach black inner-city schools, although it must be clearly emphasized that the state and district directors attempt every year to persuade them to participate. The problem would be alleviated if the competition were to become a model for assessment of history/social studies, which would be compulsory for all schools.

APPENDIX A

COMPETITION QUESTIONS WITH DISCUSSION OF STUDENT RESPONSES

The following is a sampling of student responses recorded at the Alabama sixth congressional district "We the People..." congressional hearing competition. They were chosen to illustrate both the strengths and the weaknesses of the student's responses:

Unit One: Political Philosophy

o What are the basic ideas of classical republicanism?

o How was James Madison's view of republicanism reflected in the constitution?

o In a republican government, if there is a conflict between the common welfare and an individual's rights, what considerations should be taken into account in deciding which should come first? Give examples.

The students began by saying that classical republicanism its reliance on the civic virtue of its people. Classical republicanism puts the common good above all else and without each citizen acting for the benefit of the republic, the republic is likely to fail. In attempting to define the common good, James Madison said that a government that protects each individual's natural rights is protecting the common welfare. One student said, "In republicanism the common welfare is superior to individual rights and when individual rights are violated then the government must step in and place controls on those people who are violating those rights."

Perhaps because this was the first question, the students were nervous and unable to effectively make the leap to contemporary issues. They chose as their example of the conflict between the common welfare and individual rights the issue of gun control. Their analysis of the issue however, was lacking. "Gun control represents this problem. Some people say it is a violation of their rights not to allow them to have guns, but others say that because of random violence everybody's rights must be protected and controls must be placed on guns." At no time did the students express an opinion on which way they thought the issue should go and support it.

When pressed in the follow-up question period, one student simply said, "it's a problem, but the NRA has so much power that no one can get [the laws] changed." The students' response to this question was not stellar; it shows their general lack of sophistication. As the competition moved on it seemed they became more comfortable and more willing to take a stand. One African American female student interrupted a male classmate while he was discussing affirmative action programs (he did not think they were fair or democratic) and said, "Well I disagree. Democracy has never been practiced in its true form and there need to be some corrections placed on the system."

Unit Two: History and Experience

o What basic ideas in the Declaration of Independence came from the natural rights philosophy, republicanism, and constitutionalism?

o What are the main arguments used in the Declaration of Independence to justify the American

Revolution?

o The Declaration of Independence claims that people have the right to alter or abolish their government. Under what conditions do you think people should alter or abolish their government? Explain your position.

After watching the students respond to the question about the origins of natural rights, it became clear that the students did not relate the doctrine of natural rights to the Judeo-Christian heritage of the Founders. The students were unable to explain what "natural rights" meant, and resorted to historical and narrative descriptions of what happened in the eighteenth century. We asked one of the judges, Professor William Collins who teaches political science at Samford University (the campus on which Cumberland Law School is located), why they seemed not to have heard of the Judeo-Christian tradition and its influence on the Declaration of Independence and the Constitution. He replied that for the past ten years history has not been taught in universities--and therefore in the schools--as a continuous narrative. The students don't recognize the continuity of influences, and, as we observed, don't seem to notice the need to explain where ideas come from.

Unit Three: The Constitution

o What were the major differences between the Federalists and the Anti-Federalists?

o How did the arguments of the Federalists and the Anti-Federalists reflect their viewpoints regarding natural rights, republicanism, and constitutionalism?

o What is the relevance of arguments or points made by the Federalists and Anti-Federalists to contemporary events under our present government?

In the six-minute follow-up period a judge asked, "Is the system of checks and balances currently working? Were the anti-Federalists correct in their fear of a strong Federal government. For example, in the current Gulf situation who has the right to declare war--congress or the president?"

The student replied extemporaneously, "While congress is the only body which can officially declare war, the Constitution provides the president with the power to set up the inevitability of war by ordering troops to the Gulf." This is an example of the high level of thinking and grappling with pertinent issues which the "We The People..." competition and curriculum can produce, although students do not always rise to the occasion.

Unit Four: Establishment of Government

o What is the power of judicial review and how was it established in the United States?

o Is the power of the Supreme Court to overrule the will of a congressional majority consistent with the principle of representative government? Why or why not?

o If the Supreme Court did not have this power, in what other ways could the rights of minorities be protected from the will of the majority? Take and defend a position on which way you think the rights of minorities should be protected.

The students responding to this question focused on the rights of the handicapped. They stated

that tyranny by the majority over the minority stood against the principles of representative government and that in a representative government the basic needs of all citizens should be addressed. Judicial review is one of the ways in which the will of the majority can be held in check when it impinges on the rights of minority.

In the case of the handicapped, the students postulated that while handicapped people may have physical or mental limits many can be contributing members of society, provided they are given access to buildings, jobs, and education. The perspective of the handicapped enriches our nation as long as others appreciate how they see the world and what that view can contribute to everyone.

Again we found it interesting given Birmingham's place in history that none of the classes chose the civil rights movement as an example of a minority's successful attempt to change the will of the majority.

Unit Five: Fundamental Rights

o What are the basic purposes of the Fourteenth Amendment's equal protection and due process clauses?

o How are the equal protection and due process clauses related to the principle of limited government?

o What are some specific examples of how the equal protection and due process clauses of the Fourteenth Amendment have been used to protect the basic rights of the people?

The government should remove itself from the lives of its citizens as much as possible, said the students, although some protections are needed. The equal protection and due process clauses protect citizens from the power of the State as opposed to the Federal government and ensure that no one can be deprived of their rights without "due process of law." The students implied that this puts the burden of protection and the responsibility of regulation on the State governments.

The team's examples were child labor laws and anti-trust laws. In both cases the "States had to go against established and powerful industries and basically tell them that what they were doing was not fair and that they had to stop. The state had to ensure that they had stopped."

During the follow-up questions the students were not able to effectively differentiate between substantive and procedural due process. In fact, they managed to talk around the issue until time was called.

Unit Six: Responsibilities of the Citizen

o Do all citizens have a responsibility to participate in the political life of the nation? Why or why not?

o If citizens should participate, what forms should that participation take?

o Is civil disobedience ever a justified form of political participation? Explain your position.

The students said that there are many different forms of participation in government but that all

citizens are responsible for participating on the minimal level of voting and maintaining an awareness of the issues. They also showed, using a voter poll taken in the early 1980's, that the average person did not vote and had little or no interest in voting.

APPENDIX B

METHODOLOGY

Ruth Mitchell and Amy Stempel from CBE and Kathleen Fulton from OTA visited Birmingham, Alabama on December 4 and 5, 1990, during the Alabama sixth congressional district "We The People..." congressional hearing competition. Dr. David Sink and Janice Loomis, the State and District Coordinators organized our visit so that we were able to speak with judges, teachers, students, and organizers before the competition and then observe the competition.

We attended dinner with Dr. Sink, Ms. Loomis, two judges, a teacher, and a student the night before the competition. We asked questions pertaining to the perceived value of the competition, the acceptance of the competition, and the drawbacks of the competition.

At the competition, Dr. Ruth Mitchell observed the same judges throughout the competition as they administered the same questions to the different groups. Amy Stempel observed the same class throughout the competition as they responded to each question. Kathleen Fulton observed and videoed random sessions to provide an overview of the program.

Contacts

Dr. Elaine Craig
Center for Civic Education
5146 Douglas Fir Road
Calabasas, CA 91302
(818) 340-9320

Dr. David Sink
State Coordinator, Bicentennial Competition
Department of Political Science and Public Affairs
The University of Alabama at Birmingham
UAB Station
Birmingham, AL 35294
(205) 934-9680

Janice Loomis, Executive Director
Alabama Center for Law and Civic Education
c/o Cumberland School of School
Lakeshore Drive
Birmingham, AL 35209
(205) 870-2433

Linda Jones, U.S. history teacher and state competition winner
Shades Valley Resource Learning Center
1810 25th Court, South
Birmingham, AL 35209
(205) 879-0531

Sue Ann Willis, Judge in district and state competition

Attorney-at-Law
Burr and Forman
3000 SouthTrust Tower
Birmingham, AL 35203
(205) 251-3000

Christopher Womack, Judge in state and national competition
Assistant to the Vice President for Public Affairs
Alabama Power Company
P.O. Box 2641
Birmingham, AL 35291
(205) 250-2607

APPENDIX C

National Bicentennial Competition on the Constitution and Bill of Rights

directed by the Center for Civic Education
cosponsored by the Commission on the Bicentennial of the United States Constitution
funded by an Act of Congress

Instructions For Judges

Structure of the Hearings

During mock congressional hearings on the Constitution and Bill of Rights, high school students will appear before "congressional committees" consisting of three judges from the community. As members of a committee, judges will hear oral presentations from groups of students on selected constitutional issues. Students should be treated as experts on their particular topic who have been asked to testify before the congressional committee.

The chair of the committee will conduct the proceedings. The chair may be one of the judges or someone in addition to the three judges, such as a member of Congress, the district coordinator, or a prominent community member. Judges are encouraged to play the role of a congressional committee member as much as possible. Certain procedures, such as formally addressing the students (as in "the chair recognizes the group from Jefferson High"), will add to the sense of authenticity.

Each competing class will bring six groups to the hearings, one group for each of the six units in the textbook, *We The People....* Each group has been given three questions on its topic to prepare in advance. The group will not know which of the three questions it will be required to answer until the question is asked by the judges.

The students will respond to their group's question with a four-minute prepared presentation during which they may use notes. They will then respond for six minutes to follow-up questions by the judges. Each of the six class groups will thus appear before the judges for a total of ten minutes.

The judges will listen to each group's presentation, question the group on its topic, and score each group. The class's total score will consist of the combined scores received by its six groups. The class with the highest total score will be declared the winner and will advance to the next level of hearings.

Hearing Questions

As mentioned above, the panel of three judges will select one of three questions to ask each student group. These questions have been provided to teachers in advance of the hearing by the Center for Civic Education.

After hearing a group's prepared presentation, judges will ask the students follow-up questions during a six-minute period. These follow-up questions constitute the judges'

most challenging task. The Center for Civic Education will provide suggested follow-up questions which coordinators will give to judges, and judges are encouraged to use these questions whenever appropriate. However, it is essential that judges focus much of their questioning on what students said in their four-minute presentation, keeping in mind the scoring criteria described below.

Judges should phrase follow-up questions concisely. It is best not to ask multi-part follow-up questions. If a multi-part question is used, the question should be asked one part at a time, allowing the students to respond before proceeding to the next part. During the final minute (indicated by a signal from the official timer), questions should either be brief summary questions or requests for concluding remarks.

Timing

For each hearing, there will be an official timer, who shall be referred to as the "Committee Clerk." The clerk should not be one of the judges. Clerks should keep strictly to the framework of ten minutes per team: four minutes for the initial oral presentation and six minutes for the follow-up questioning. Clerks will silently notify groups by holding up a card when they have one minute left in their initial oral presentation time and will silently notify judges when there is one minute left in the follow-up question period. When time has expired, the clerk will stop all participants by saying "Time!"

Every effort should be made to observe the time limits. However, at the discretion of the judges, extra time may be allowed a group to compensate for any situation that significantly reduces the time allowed for response to questions. For example, if a judge has taken an inordinate amount of time to phrase a question, extra time may be granted to the student group for its response.

Use of Notes or Materials

Students may use notes on paper not larger than 8 1/2 X 11 inches during the four-minute presentation period. No other materials or references may be used. Neither notes nor any other materials may be used during the follow-up question period.

Judging Criteria

Judges will score each student group on the basis of six criteria. These criteria are: understanding, constitutional application, reasoning, supporting evidence, responsiveness, and participation. These criteria are more fully explained on the attached "Congressional Hearing Group Score Sheet." Throughout their assessment of student performance, judges should be careful to base their scores on the six established criteria and not on whether or not the judges agree with a position students might take.

For each criterion listed, each judge will score the group on a scale of 1 to 10, with 10 being the best score. Judges must use a separate form for each group. Before beginning a hearing, judges should ascertain that they have the appropriate score sheet (correct teacher name, school name, and unit number). The scorekeeper will total class scores on

a separate summary form. The class with the highest total score will be the overall winner.

The scoresheets may be retained for all groups seen during one session and may be revised during that period. For example, if, after judging 2 or 3 groups, the judges think they were too harsh or too easy on the first group, they may adjust their scores. However, once the scores have been turned in, it will not be possible to revise scores. Judges may consult with one another about scoring; however, they do not need to agree on their scores.

Special care must be taken in judging a group on the sixth criterion, participation. Groups in which most members speak should be rated higher than those in which one or two "star" students dominate the entire 10 minutes. It is acceptable that one or two students make the opening statement, but the other students should attempt to answer the follow-up questions. Since it is recognized that certain students have an especially difficult time with speaking in public, it is the policy of the Bicentennial Competition that a group not be penalized for containing one or two students who do not participate because of shyness, language problems, or handicaps, etc. In rating participation, judges should evaluate the extent to which most students respond.

Judges will also be asked to award up to 100 bonus points to each class group for its overall performance above and beyond the six basic criteria. In the event of a tie, these bonus points for a class's six groups will be totaled and the tying class with the highest total number of bonus points will be declared the winner. Tie-breaker points should be awarded for overall performance of the group using the following guidelines:

Excellent	90 to 100 points
Very Good	80 to 89 points
Average	70 to 79 points
Below Average	0 to 69 points

Feedback

Although students should not be told how their presentations were scored, they should be given some feedback. Because of the time constraints, that feedback will have to be brief. It should include comments on the group's strengths and suggestions for improvement.

Conflicts and Protests

In the event of a protest or conflict during the course of the competition, final arbitration shall be performed by the district coordinator or designee at the congressional district level, and the state coordinator or designee at the state level, subject to the rules and regulations of the National Bicentennial Competition on the Constitution and Bill of Rights contained in *Culminating Activities Guidelines and Rules* and in the most recent guidelines in the National Bicentennial Competition's *Administrative Handbook*.

Sample lesson
from the
"We The People..." textbook

Lesson 1

What would life be like in a "state of nature"?

Purpose of Lesson 1

The first and second lessons in this unit will introduce you to some of the basic ideas of the natural rights philosophy which were of great importance in the development of our government. The most important ideas contained in Lesson 1 are listed below.

When you finish this lesson, you should be able to explain these ideas, their place in the natural rights philosophy, the source of the ideas, and how they influenced the kind of government the Founders created.

political philosophy
state of nature
law of nature
consent
natural rights philosophy
human nature

Library of Congress

John Locke, English political philosopher, 1632-1704

The natural rights philosophy

We hold these Truths to be self-evident, that all Men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the Pursuit of Happiness--That to secure these Rights, Governments are instituted among Men, deriving their just Powers from the Consent of the Governed, that whenever any Form of Government becomes destructive of these Ends, it is the Right of the People to alter or to abolish it, and to institute new Government...
(Declaration of Independence, *1776)*

This excerpt from the Declaration of Independence includes some of the most important philosophical ideas underlying our form of government. They are ideas that had been accepted by almost everyone in the American colonies long before the Revolutionary War. They had been preached in churches, written in pamphlets, and debated in public and private. These basic ideas had been developed and refined by political philosophers such as the Englishman John Locke (1632-1704) and by many others in Europe and in the colonies. Of these philosophers, John Locke was the most important influence on the thinking of the Founders at the time of the Revolution. The political philosophy Locke wrote about is often called the natural rights philosophy.

The natural rights philosophy is based on imagining what life would be like if there were no government. Locke and others called this imaginary situation a state of nature. Whether or not such a state ever existed, thinking about what life would be like if there were no government was very useful to philosophers such as Locke in answering the following questions.

1. What is human nature, that is, what traits of personality and character, if any, do all people have in common? For example, are all people basically selfish or do they tend to care for the welfare of others?

2. What should be the purpose of government?

3. How do the people running a government get the right to govern?

4. How should a government be organized?

5. What kinds of governments should be respected and supported?

6. What kinds of governments should be resisted and fought?

John Locke's answers to these questions were the answers accepted by most of the Founders. They used these ideas to explain and justify their declaration of independence from Great Britain. They also used these ideas in writing the various state constitutions after the Revolutionary War and later in writing the Constitution of the United States.

Problem solving

It's your turn to think like a philosopher

In order to understand the natural rights philosophy, it is helpful to try to answer the questions it deals with. Some of the most important of these questions are included in the following exercise. You may not all agree on your answers to these questions.

It is important to know that you are not alone. At various times in history, people have had very different views on these matters.

Imagine that all of the students in your school were transported to a place where there were enough natural resources for you to live on, but where no one had lived before. When you arrived, you had no means of communicating with people in other parts of the world. On the basis of this situation, answer the following questions, discuss your answers, and then compare your answers with those of John Locke which follow.

1. Would there be any government or laws to control how you lived, what rights or freedoms you exercised, or what property you had? Why?

2. Would anyone have the right to govern you? Would you have the right to govern anyone else? Why?

3. Would you have any rights? What would they be? Would it make any difference if you were a man or a woman?

4. What might people who were stronger than others try to do? Why?

5. What might the weaker people try to do? Why?

6. What might life be like for everyone?

Compare your answers with John Locke's

Your answers to the above questions may be similar to the following answers developed by John Locke or they may differ. In this lesson, however, we are focusing on understanding Locke's answers because they were widely shared by Americans living during the 1700s and played a very important role in the development of our government.

1. It is clear that Locke believed that there were laws in a state of nature. He said, "The state of nature has a law of nature to govern it which obliges every one .. no one ought to harm another in his life, health, liberty, or possessions . . ." These laws were "the Laws of Nature and of Nature's God" as Thomas Jefferson called them in the Declaration of Independence.

However, the problem in the state of nature would be that you and others would probably disagree on what the "laws of nature" are and there would be no one with the right to help you settle your disagreements. This is because there would be no government to say what the law was or to enforce it.

According to Locke, there wouldn't be any government because a government can't exist until it has been created. And a legitimate or just government cannot exist until the people have given their consent to be ruled by it. Thomas Jefferson included this idea in the Declaration when he wrote that

"Governments are instituted among men, deriving their just powers from the consent of the governed; . . ."

2. No one would have the right to govern you, nor would you have the right to govern anyone else. According to Locke, the only way anyone gets the right to govern anyone is if the person to be governed gives his or her consent. If people haven't consented to the creation of a government, there is no government.

3. You would have the right to life, liberty, and property. This means you would have the right to defend these rights if other people threatened to take them away. These were called natural rights. We now call them fundamental, basic, or human rights.

4. Locke believed that people are basically reasonable and sociable, but that they are also self-interested. Since the only security people would have for the protection of their natural rights would be their own strength or cunning, people who were stronger or more able would often try to take away the life, liberty, or property of the weak.

5. Weaker people might try to protect themselves by joining against the strong.

6. Because there would be no laws that everybody agreed upon and no government to enforce them, everybody's rights would be very insecure.

The problems of a state of nature

John Locke and most Americans in the eighteenth century thought that people are equal in their rights to life, liberty, and property. They have these natural rights just because they are human beings. This means that in a state of nature, no one would have the right to interfere with your life or property or your freedom to do as you wished. However, Locke also believed that in that kind of situation, because of human nature, there would always be people who would try to violate your rights. Since there would not be any government, you and others would have to defend your rights against those who might want to take them away. The result would be that in the state of nature, your rights would be insecure and you would be in constant danger of losing them.

For Locke and those who shared his ideas, the great problem was to find a way to protect our natural rights so that we can live at peace with one another and enjoy them.

Reviewing and using the lesson

1. The Declaration of Independence contains some of the most important ideas of the American system of government. These are often called the natural rights philosophy. Briefly describe some of the major points of this philosophy.

2. According to Locke, what is the purpose of government?

3. It has been said that since people are not equal in their intelligence and character, it is unjust for all to have the same rights. Do you agree? Why?

4. Imagine yourself living in a community where all governmental authority has broken down. Violent lawlessness is widespread. Do you think any government is better than none? Why?

5. How might your ideas about what human beings are "really like" influence your ideas about what type of government is best?

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

SUMMARY AND ANALYSIS

Section I. Introduction

Performance assessment is in its infancy as a technology for educational evaluation. Although it shows considerable promise for assisting in the reform of education, it is by no means a panacea. It presents problems of a different kind than those posed by norm-referenced multiple-choice tests; for example, the validity of the assessments and the reliability of scoring techniques have not been researched fully yet. But since performance assessments can influence curriculum and instruction by modeling desired practices, their potential is too important for them to be dismissed as psychometrically naive or as too costly to be considered seriously.

Our six cases illustrate both the potential and the accomplishment of performance assessment, for they range from the College Board's Advanced Placement Studio Art Portfolio Assessment to a home-grown school-level assessment in which parents interview elementary students. Three of the cases--the Arizona Student Assessment Program, the South Brunswick K-2 Portfolio, and the California Assesment Program in relation to the San Diego City Schools--exemplify the use of performance assessments to influence teaching. Another, the Advanced Placement Studio Art Portfolio Assessment has been in place for almost 20 years, growing each year. In contrast, the Arizona Student Assessment Program has just been adopted by the state and will not be officially scored until June 1992. One of the cases, the "We the People..." competition sponsored by Center for Civic Education, was chosen because it offers a model of social studies assessment in the form of a competition designed to increase knowledge of the U.S. Constitution.

An important feature common to all--even the Advanced Placement Studio Art Portfolio Assessment--is a close connection between instruction and assessment. In many conversations and interviews, it seemed that the program and the assessment were identified in respondents' minds. The assessment was seen as a natural consequence, even part of, instruction. This contrasts with the intended separation of norm-referenced multiple-choice tests from instruction.

Another common feature is the influence of purpose on most aspects of the assessments--the choice of the assessment's form; the experience of implementing it; the role of the teachers; and the budgets which share the cost (for example, if an assessment is intended to steer instruction in a certain direction, then some of the costs can be charged to teachers' professional development). More than one purpose can be intended, and assessments can also be used for purposes not

intended. The qualities and the shortcomings of the assessments can be understood only in light of their purpose.

This summary and analysis follows the organization of each case study into eight sections. It attempts to draw generalizations from this obviously disparate set of cases by pointing to common and contrasting elements. The summary and analysis is intended, like the case-studies themselves, to add to the public store of knowledge about educational assessment.

Section II. The purposes of the assessments and the rationale for choosing them

As mentioned in the introduction to these case-studies, educational assessment generally serves three purposes: accountability to legislatures and funding bodies; feedback on instruction; and student selection and sorting (certification). In addition, performance assessment can fulfill a purpose--modeling desired curriculum and instruction--that norm-referenced multiple-choice tests cannot.

What emerges from our case-studies is that performance assessments can fulfill two purposes at once, and may be consciously intended to do so. The Arizona Student Assessment Program (Arizona Student Assessment Program) is the clearest example of an assessment designed to perform two roles. Its results will be reported to the Arizona legislature to indicate progress in implementing the curriculum frameworks, the Essential Skills Documents. But the form of the assessments was deliberately chosen to move curriculum and instruction in the direction of implementing the frameworks and showing teachers the desired approaches to the teaching of reading, writing and mathematics. While our case study of the relationship between the California Assessment Program (California Assessment Program) and San Diego City Schools is not focused on the assessments as such, we should point out that the California Assessment Program statewide assessments originally had the same two purposes as those of the Arizona Student Assessment Program. The California Assessment Program assessments were the predecessors and to some degree the models for the Arizona Student Assessment Program (as well as for assessments in other states). However, now California Assessment Program, with the assistance of the San Diego City Schools and other school districts in California, is seeking to design a statewide assessment which will accomplish three purposes at once: accountability; modeling desired instruction; and providing individual student scores for certification, selection, and sorting.

These are examples of conscious intention to serve more than one purpose. However, it seems inevitable that purposes are blurred in practice. For example, the Advanced Placement Studio Art Portfolio Assessment is our clearest example of an assessment designed to certify individual student achievement. Yet it influences instruction, as we have shown from interviews with teachers, and its results, together with all the Advanced Placement scores, are used as measures of high school quality, despite repeated protests by the Educational Testing Service that these examinations are not suitable for school accountability.

The case is similar with the "We the People..." competition. The assessment is intended as selection and sorting--but of groups of students, not individuals. As an extracurricular competition, "We the People..." has few consequences for the students' educational future. Nevertheless, it influences instruction and it also reveals important information about the teaching of history/social studies.

Using assessments intended for one purpose to serve another is of course not confined to performance assessments. Multiple-choice norm-referenced tests are also used for purposes not

intended--in fact, a major reason for the development of performance assessments was the perception that tests intended for accountability or for student certification were influencing curriculum and instruction in ways not considered desirable. The issue is whether the extension of purpose is a use or a misuse, that is, whether it supports good instruction or subverts it. For example, in the Longfellow Art Assessment in Bozeman, Montana, Ray Campeau, its designer, does not want to report the results in detail to teachers. He fears that they would begin to emphasize aspects of the curriculum that apparently were learned less well than others, instead of teaching all of the curriculum equally.

The lesson from analyzing the purposes of our six case-studies is a need to anticipate at the outset what purposes the assessment might be used for beyond the stated and intended reasons. Attempts to control use of assessments by earnest statements about their purposes do not work, as the College Board knows to its chagrin, not only from the misuse of Advanced Placement courses and scores as indicators of high school quality, but also from the widespread use of the Scholastic Aptitude Test (SAT) results to judge schools and school districts. The public needs to be helped to understand the purposes of assessment, especially when its form changes from the familiar multiple-choice to performance assessment, but the designers of assessments need to strategize carefully so that assessments cannot be used for purposes detrimental to the overall aims of education.

The rationale for choosing specific forms of assessment is related directly to the purpose and to the nature of what is to be assessed. This is an important difference from norm-referenced multiple-choice tests, where there is no choice of form.

The fit between the subject assessed and the form of the assessment is tightest in the case of the Advanced Placement Studio Art Portfolio Assessment. Looking at examples of production seems the only authentic way to assess accomplishment in studio art. The exigencies of a mailable portfolio restrict what can be offered as evidence of ability, but practical considerations preclude other possible methods of review, such as panels of reviewers traveling to school sites.

It is also hard to imagine a more appropriate way of assessing the literacy development of young children than the interviews and observations of the South Brunswick K-2 portfolio assessment. The fit is perhaps loosest in the case of "We the People..." where the form of assessment was chosen for reasons external to the material. It would be perfectly possible to assess understanding of constitutional issues--even in competitive circumstances--by means of written responses to questions; individual oral presentations; or recreations of famous constitutional crises in the form of trials, perhaps videotaped. But these alternatives did not meet the need for an assessment that would take a short time to complete, would involve all members of a school class, and would involve the public.

The Longfellow Elementary Art Assessment in Bozeman MT is also a loose fit with its subject (learning about art), since there are other ways in which the knowledge could be assessed. But given the age of the students involved, which makes writing a less practical alternative, direct interviews have merit, although they bring with them a number of so far unresolved problems.

In general, choosing an appropriate performance assessment is a more complicated issue than the purchase of a multiple-choice test. It requires balancing a number of factors: purpose, the nature of the subject to be assessed, the time available, and the degree of rigor needed in scoring. We cannot say that some of our cases are more successful than others in choosing a form of assessment, since these factors are not comparable across the cases.

Section III. Experience to date

In general, the experience of setting up performance assessments is likely to be more successful and present fewer problems where the assessment is supported by and embedded in an established educational system than if it is either added on for a specific purpose or remains outside the system.

By far the longest running and most accepted of our cases is the Advanced Placement Studio Art Portfolio Assessment. It has experienced few problems because it is part of an established national program, the College Board's Advanced Placement program, and enjoys the benefits of being associated with it.

In contrast, the Arizona Student Assessment Program has no track record at all as an implemented assessment. But it has been established by the state's top educational administration, with the support of the Arizona legislature. The conference we attended for our site visit was the first occasion on which the Arizona Student Assessment Program was introduced to the vast majority of the state's districts, except for those who had taken part in pilot tests. The Arizona Student Assessment Program process has clearly been "top-down" so far. So too has the South Brunswick K-2 Portfolio assessment, although teachers claim ownership because the district administration has involved them from the beginning. But the portfolio's use is mandated by the administration, and it is clear from the teachers' own remarks that it originated from the administration's desire to support developmentally appropriate teaching with similar assessment. The California Assessment Program also proceeded top-down at first, and is now in the process of equalizing its relationship with the San Diego City Schools (and other large California districts) in order to produce an integrated statewide assessment system. The point to be noticed is that "top-down" does not mean dictatorial and does not necessarily imply that teachers have been ignored and excluded from ownership.

The Longfellow Elementary School Art Assessment arises from a different place in the school district organization. Its designer is not in a leadership position in Longfellow School, although he enjoys considerable prestige in the Bozeman district as a whole. Because he is not an official leader, he cannot use Longfellow School's resources freely in the administration of the assessment. He is dependent on uncertain volunteer help, which limits the rigor of the assessment.

The Longfellow assessment is outside the school in an odd parallel with the relationship with test publishers: the art assessment was designed because the Longfellow School principal felt the need for an evaluation of the program and could not find or afford a published test, so it replaces a test in the minds and attitudes of the school personnel. The teachers are not part of the process, because the designers of the assessment did not want to increase the burdens already placed on teachers. The assessment was designed in consultation with the principal and a committee of three teachers, just as a test would be chosen by the principal and a few teachers. The parents fill the role of external scorers, somewhat ironically, and the results were expected by the teachers with some of the same anxiety prompted by test scores. Some (although by no means all) of the problems experienced--unreliable volunteer help, resentment among parents asked to volunteer again when they thought they had done enough, resentment among teachers who didn't receive the results of the tests as soon as they would have liked them--are attributable to treating the situation as if it were the same as applying a published test, but without the organization supporting a test publisher.

"We the People..." is entirely outside the school system, but its success is not dependent on it. It is institutionally secure in its own organization, the Center for Civic Education, and is supported by a

prestigious national commission. Two major problems remain: the need for further funding after the Bicenntennial Commission's grant expires in 1992; and the possible shape of an ongoing relationship with the school system, with or without external funding. The solutions to both are outside the scope of this study.

Another general conclusion that can be drawn from the experience of some (although not all six) of our cases is a changing relationship with test publishers. We chose the Arizona Student Assessment Program because we wanted to see how a state and a publisher cooperated to design and produce performance assessments. The relationship between test publishers and their customers is usually based on marketing--the test publisher sells the test on the basis of its ability to meet the district or state's assessment needs. In Arizona, the Riverside Publishing Company responded to the state department's request for proposals with a program based on the state's Essential Skills Documents and then worked with the state department on the various forms of the assessment. In California, the California Assessment Program contracts with a test publisher (formely the Educational Testing Service, and now Advanced Systems in Measurement and Evaluation, Inc. of Dover NH) to administer the scoring of the writing assessments; similarly, San Diego City Schools is working with Psychological Corporation to design their California Framework Focused Examinations, intended to reinforce the teaching of the California frameworks at the non-California Assessment Program grade levels. Since our site visit, South Brunswick school district has begun a partnership with researchers from the Educational Testing Service to develop a scoring scale for the K-2 portfolio assessment.

This pattern of cooperation with test publishing is beginning to spread: a similar though not identical relationship is operating in Maryland, between the Maryland Department of Education and CTB/McGraw Hill, and in Vermont between the designers of the statewide portfolio assessment and Advanced Systems in Measurement and Evaluation, Inc. This pattern of relationships should be mutually beneficial, since the school district gets the benefit of expert and experienced advice, while making sure that the assessments serve its purposes, and the test publishers get business.

Finally, the experience of some (again, not all) of the six cases has been affected by the strength of the personalities involved. It is clear that there would have been no Longfellow Elementary School Art Assessment without the presence and forceful personality of Ray Campeau; the same is true of South Brunswick, where Willa Spicer's leadership is acknowledged to be an essential ingredient in the development of the K-2 portfolio assessment.

The presence of forceful personalities is fortuitous, an ingredient in the political and social background which affects the development of assessments as it affects all other parts of the educational system. These studies cannot be understood outside the context provided by the politics and personalities involved. The content of Arizona's assessments has been affected by special interest groups, for example.

Another example: the California Assessment Program was terminated in 1990 because of a political and personal conflict between the governor of California and the State Superintendent of Public Instruction. But this apparent disaster had a silver lining: the hiatus in testing has provided an opportunity for the design of an integrated system of educational assessment in California, to include the state and school districts as equal partners, and to provide student scores, as well as program monitoring and instructional modeling, from performance assessments. It may become the first example of a moderated system in the U.S.

Section IV. Teachers' involvement in performance assessment

As we mentioned at the beginning of this summary and analysis, the purpose of the assessment affects every aspect of it. The involvement of teachers varies in degree accordingly: where the purpose is to assess the effectiveness of a program, as in the case of the Longfellow Elementary School Art Assessment, the teachers need not be involved; where the intention is to change instruction, as it is in South Brunswick, Arizona, and California, teachers are brought into the process; and where selection and sorting are intended, as in the Advanced Placement Studio Art portfolio assessment and the "We the People..." competition, teachers are outside the assessment process but act as coaches for their students. Their coaching is an intensification of regular teaching, not, as it is in coaching for norm-referenced multiple-choice tests, an activity which teachers describe as different from normal instruction.

In none of our cases were teachers the leaders in developing performance assessments. Where they were actively involved with either development or scoring, they worked under experts or administrators. The California Assessment Program writing assessments were developed by teachers under the direction of Professor Charles Cooper of the University of California San Diego and California Assessment Program consultants. They score the assessments after rigorous training sessions conducted by experts from Advanced Systems in Measurement and Evaluation Inc. The same pattern is being followed with the mathematics performance assessment, and with the proposed assessments in science and history/social studies.

In Arizona, teachers were involved in pilot testing the assessments in spring 1990, including the scoring, but have been less involved in the development of assessments than they were in California. Large numbers of teachers are now being trained in scoring conferences (taking place after our site visit in November 1990) by the Center for Educational Assessment, (a subcontractor of the Riverside Publishing Company,) so that they can score the field-tests of the assessments and be ready to score all the assessments when they are formally implemented in 1992.

Teachers in South Brunswick have revised the portfolio for two summers and are likely to continue. They feel that they are at the center of the process, although it was initiated by an administrator.

It seems as if supported participation is the most comfortable position for teachers in the assessment process: they need to be involved and feel a sense of ownership, but they contribute most to the process of development and scoring when experts guide them. Their training usually does not include much information about or training in educational assessment, so that they are not confident without expert guidance.

The impact on teachers where instructional change is the purpose of the assessment is clear in the case of the California Assessment Program's writing assessment and in the case of South Brunswick. There is research to document that the writing assessment increased the amount of writing in California junior high schools and that teachers understood the need to assign writing in different genres. South Brunswick teachers told us directly that the portfolio assessment was an invaluable tool for them in adapting their instruction to the needs of their students both collectively and individually. There is no evidence of the effects of the Arizona Student Assessment Program on teachers yet, although the teachers we interviewed thought that it would cause changes in the way they taught.

Section V. Scoring

There is a wide variation in scoring methods in our cases, ranging from the psychometrically strong scoring of the Advanced Placement Studio Art Portfolio Assessment to the complete lack of scoring in the South Brunswick K-2 portfolio assessment. With the exception of the well established programs, (the Advanced Placement Studio Art and the California Assessment Program's writing and mathematics assessment), there is more faith than expertise in scoring.

The established programs use group grading, in which standards are established by a consensus process and then applied by "readers" to student products (the art portfolios, essays, or written answers). The consensus or standard-setting process is the most important part of the procedure, and can take up to one-third or a half of the allotted time. All the readers or scorers look at examples of the work and discuss their criteria until they agree on grades for each example. In the Advanced Placement Studio Art Portfolio Assessment, products receive scores from more than one "reader," and must be scored by another reader or the leader if the scores differ by more than a certain number of points. The single score that a student receives for Advanced Placement Studio Art is based on scores for three different parts of the portfolio, weighted and scaled appropriately.

In the California Assessment Program writing assessments, essays and answers are read by a single reader, but various techniques are employed to maintain consistency of grading: marked papers already read are circulated back into the pile to see if they get the same grade again; the table-leaders randomly reread papers to make sure that readers are consistent; examples of graded papers are kept available for comparison as "anchors." Using these techniques, the interrater reliability for the California Assessment Program writing assessments is about 90 percent in a single year. It is not so high across years, as determined by rescoring papers from one year to another (some prompts are kept from year to year to make this possible). This remains an unsolved problem for California Assessment Program and for other states and districts using group grading if they want to make longitudinal comparisons.

The Educational Testing Service builds trends into its scoring for the Advanced Placement Studio Art Portfolio Assessment by comparing the scaled scores to the previous two years' scores. There is also continuity of personnel, so that a chief reader can defend a dramatic difference (which almost never happens) between one year and the next from personal experience and the data.

The Arizona Student Assessment Program has not been formally scored yet. Its scoring will be conducted by the Center for Educational Assessment, subcontracted by the Riverside Publishing Company, which will train the teachers and organize the scoring procedures using much the same techniques as Advanced Systems in Measurement and Evaluation Inc. does for California Assessment Program. However, as we made clear in our case-study, there are so far unresolved problems presented by assessments that extend over two days and by the necessity that an entire class complete a single assessment at once in order to accommodate discussion.

In these three cases, (Advanced Placement Studio Art, California Assessment Program, and Arizona Student Assessment Program) psychometric issues are the business of companies which have long experience with them. Because the California Assessment Program and Arizona Student Assessment Program focus on the form of the assessments and their role as models of instruction in their communication with teachers and school districts, they may give the impression that they are less concerned with validity and reliability. The impression is false: they take expert advice from their contractors and technical advisory groups.

Since scoring is not an issue for South Brunswick yet (the development of a scale in conjunction with Educational Testing Service researchers began after our case-study was completed), the only other cases which use scores are the Longfellow Elementary School Art Assessment and "We the People..." There is no point in pretending that either of these is interested in psychometric sophistication. They both chose the simplest method available to get comparative scores. In each case, judges assign scores which are then totaled--no weighting, no scaling, just addition.

"We the People..." uses the simplest possible kind of scoring: judges give individual scores, with or without consultation with colleagues or adjustment by comparing later performances with the ones they judged first. The Longfellow Elementary School Art Assessment is a little more sophisticated, since it does include some discussion about the number of points to be given for evidence of more or less understanding of art, and it is conducted in a group, so that discussion about ambiguous scores is possible. The group scoring activity is based on Ray Campeau's experience as a "reader" for the Advanced Placement Studio Art Portfolio Evaluation, which gave him a model for scoring the responses collected by parents in interviews with the children.

Neither of these scoring schemes would stand up to a moment's scrutiny from a psychometrician--but they don't need to. The purposes of each assessment are served adequately. The Longfellow Elementary School Art Assessment needs only to show that the Art in Action curriculum has produced growth in knowledge about art as demonstrated by ability to respond to pictures. The "We the People..." competition plays for high stakes only at the culminating national level, and even then the academic fate of individual students does not hang on the results.

Of course if either of these programs were to be incorporated into school systems where information of a high degree of accuracy was required, the scoring schemes would have to be made rigorous.

Section VI. Costs

In December 1990, the Council of Chief State School Officers compiled a chart of performance assessments in the states so that comparisons could be made easily. One of the columns was labeled "Costs." As the data came in, the numbers fluctuated wildly because different respondents thought of costs differently: some reported costs of development ("$2 million" in one case), some costs of administration ("$5 a student"), and some combined them. Finally it was decided to eliminate the column because it could provide no meaningful information and would require explanation no matter what it included.

The major problem in approaching the costs of performance assessment is lack of a common base for the information. This applies both when comparing the costs of norm-referenced multiple-choice tests and when comparing the costs of different forms of performance assessment. We will first illustrate the problem by looking at the financial background of each of our cases, and then sketch out some directions in which research might usefully go to make comparisons possible and useful.

In South Brunswick, there is no point in comparing the costs of norm-referenced multiple-choice tests and the K-2 portfolio assessment because they are not in competition for funds. The norm-referenced multiple-choice tests are mandated by the state of New Jersey and there are no plans to replace them with the portfolio. The costs of the portfolio include: Willa Spicer's time as director of curriculum and instruction for the district; compensating teachers for their planning time during summers; compensating teachers who give workshops for other teachers during the year; the costs

of sending one teacher to New York University for training in Reading Recovery techniques; and paying for aides and other adult help in K-2 classrooms. Some of these costs have been met by astute juggling of available funds, including recouping the costs of training the teacher in New York by charging fees for her to give workshops in neighboring districts. The total outlays in dollar terms are not large. But a strict measure of what it costs to conduct the K-2 portfolio assessment would include all the salaries of all the K-2 teachers and aides using the portfolio, since by definition they are using it as part of their teaching.

The Arizona department of education has told the legislature that the costs of the Arizona Student Assessment Program, which includes one-third of the Iowa Test of Basic Skills each fall and the performance assessments each spring, will equal what the state formerly spent for the complete battery of tests purchased from the Riverside Publishing Company. This is intended to include the costs of compensating teachers for scoring, the costs of contracts with Riverside Publishing Company and the Center for Educational Assessment in Columbia MO for running scoring sessions, and the costs of the conferences to introduce the program and its scoring. But it does not include the time spent not only by the state administrators, but also by the administrators at the district and school level on decisions about the assessments (including how to use the materials given to each district at the October 1990 conference), nor by the teachers. As in the case of South Brunswick, if the teachers use the assessment material for teaching (as they are encouraged to do), then their time could arguably be regarded as a cost of assessment.

The situation is a little clearer in the case of California, where there is a budget for the California Assessment Program. However, this pays for all California Assessment Program's testing activities, which include multiple-choice tests as well as performance assessments, and the development of the Golden State Examinations, end-of-course examinations something like the New York State Regents' Examinations. The California Assessment Program budget also pays for its consultants to travel throughout the state to work with groups of teachers developing assessments, and for the teachers' time. It also compensates teachers for their time when they meet for group grading sessions and for the professional services of Advanced Systems in Measurement and Evaluation Inc., besides other consultants and services.

One issue with the California Assessment Program budget is separating out the costs of performance assessments as opposed to that of other tests. Another issue is the costs of working with teachers on development and scoring. Since this is regarded as excellent professional development, an annual percentage of the costs of scoring the California Assesment Program assessments could be charged to the state budget for that purpose--or charged to the districts who benefit from the teachers' experience. Similarly, some of the work done in these sessions is also curriculum development, since teachers learn what should be included in and what excluded from district curriculum. An annual percentage could be charged to a district or state budgets for curriculum development.

We were given figures (quoted in the case-study) for the San Diego City Schools tests, but they did not include the costs of the district office which runs testing and measurement. They also did not include teachers' time for administering both multiple-choice and performance assessments in the classroom.

The Advanced Placement Studio Art Portfolio Assessment at first seems simple to budget: $65 per student, which is an amount worked out by the Educational Testing Service and the College Board to contribute to the total cost of the Advanced Placement Program, now paying for itself after many years in the red. However, this does not include the costs of the teachers' time spent coaching

students, often outside normal working hours; or the costs of materials, often borne by the students themselves.

The Longfellow Elementary Art Assessment Program also costs more than appears on the surface. Ray Campeau was given half-day released time from his position at the high school to set up the program and the assessment (which took about four weeks of the total time), but presumably that does not include substitute time during the school year. The parents' time spent interviewing the students and scoring the responses was donated, but if the program were replicated using teachers or other non-volunteers, the costs of their time would probably be the largest part of the budget. The costs of the time spent by the principal and the three teachers designing the assessment with Campeau would also have to be calculated for a realistic sense of what the art assessment actually cost. The $2,000 spent for art materials is not part of the assessment budget, since they were not used for the assessment.

The "We the People..." competition looks like a free ride for the schools involved: no costs for the materials or for entering the competition. But the teachers attend a summer workshop, for which they are compensated by earning credits for increased pay; how much of what they learn during the workshop applies only to the curriculum and how much to the competition would be almost impossible to work out, although it is possible to calculate the increase in their salaries as a result of earning units. Teachers also spend time coaching students, sometimes outside regular school hours. If students need to travel to the higher levels of the competition, the district coordinator (sometimes with the help of the state coordinator) must secure financial support from corporate sponsors and sometimes organize fundraising efforts among students. The time the district coordinator spends on this and on the organization of the competition in general is now donated by the organization he or she works for. In addition, the "We the People..." curriculum and competition also function as professional development for those teachers who participate in addition to the formal training involved.

Performance assessment interacts closely with instruction. Its proponents argue that the model of instruction assumed by performance assessment make assessment and instruction practically interchangeable. This method of instruction, however, is new to teachers and many must re-learn to teach. Any discussion of the costs of performance assessments needs to bear these differences in mind.

Given this variety of funding situations and the omnipresence of hidden costs, how can we compare either different forms of performance assessment or different kinds of assessments? Here are some possible guidelines:

1. <u>Define clearly what is included in "assessment."</u>
   It matters less what specific definitions are adopted than that they clearly provide a base for consistent comparisons. For example, the costs of norm-referenced multiple-choice tests are frequently given as the costs of the contract with the publisher, without including the costs of the teachers' time preparing students or administrators' time negotiating with test publishers and ensuring distribution of the tests and so on. If the costs are defined as those of purchasing tests, then comparison to performance assessments must be confined to that issue. If however, teachers' time is included in the costs of performance assessment, then it must be included also for norm-referenced multiple-choice tests.

2. <u>Include costs of all activities that impact on the assessment.</u>
   The guiding question here is: if the assessment type or system were removed, what would

remain? For example (and only for the sake of argument), if Arizona cut out its new statewide performance assessment rigorously, then an entire office would disappear from the Arizona department of education, and a number of teachers across the state would find themselves with shortened hours (or, at the elementary level, out altogether), because they are teaching from the assessment materials.

3. Separate out those items which could be applied to other budgets.
To continue the example under #2, obviously the teacher using the Arizona assessment materials as curriculum is doing two things at once. When the teachers score the assessments, as they do in California for the California Assessment Program writing and mathematics assessment, they are performing a service for the California Assessment Program but also getting the benefits of the "best staff development I've ever had," as many teachers say. Thus the costs of teachers' involvement in assessment could also be charged to professional development and instructional budgets. They should only be attributed in this way, however, if the same procedure is applied to the other kinds of assessment to be compared. The important point is to recognize that some activities connected with assessment fulfill more functions than assessment alone. It would be a matter of policy to decide which budget should carry the costs.

It is beyond the scope of this study to work out the details of the model in dollars, full-time equivalents, or percentages of them. We have pointed out the inconsistencies in budgeting among our six studies that make impossible any direct comparisons among the figures we have used.

The general conclusion to be drawn is that statements about the expense of performance assessments are unfounded without the analysis we have recommended. The cost comparisons between norm-referenced multiple-choice tests and performance assessments thrown out in speeches and in answers to press questions are simply guesses at this point.

Finally, another problem with the costs of assessment must be mentioned: assessment is underfunded across the board. Dr. Dale Carlson, director of the California Assessment Program, has called for the "one percent solution," pointing out that such an important function deserves one percent of the budget--twenty times more than the California Assessment Program's funding. The Educational Testing Service's Trustees' Public Accountability Report (Educational Testing Service 1990 Annual Report) cites a "state with a strong commitment to educational assessment" which found that moving to performance assessment would cost ten times the amount that the present program costs (p. 6). Apart from the issue of comparison discussed above, even ten times as much is only half what Carlson thinks is a reasonable amount for good statewide assessment.

Section VII. Impacts on minority groups

Impacts on minority groups are a major cause of apprehension about the change from multiple-choice testing to performance assessments. Our six cases do not shed much light on minority achievement, but clearly point to different perceptions on the part of minority groups.

Arizona's minority, Native Americans, mostly Navajos, see performance assessments as an opportunity for their educational achievements to be recognized, since they believe that multiple-choice tests seriously misrepresent their abilities.

African Americans, as represented by Dr. Shirley Weber, president of the San Diego City Schools

Board, fear that performance assessments are scored subjectively and that their children will suffer from the expectations of "racist" teachers.

Only experience of both performance assessments and the curricula they need will provide information on the accuracy of these speculations. At least three years of experience will be necessary before definitive results will emerge.

Fairness to minority populations is a major issue in testing. One organization, the National Center on Fair and Open Testing (FairTest) is entirely devoted to monitoring tests, especially intelligence and aptitude tests such as the Scholastic Aptitude Test (SAT), which it perceives as unfair to some groups, including women. Fairness is especially an issue where assessment is used for sorting and selecting students.

But only one of our six cases, the Advanced Placement Studio Art Portfolio Assessment, is concerned with student certification. As far as we could discover, there have been no challenges to the fairness of the examination. The issue affecting students is not connected with the portfolio assessment itself but with colleges' refusal to grant exemption from first-year courses for students with high scores.

Of the other five cases, only the Arizona department of education has faced serious challenges from a minority, the Christian fundamentalists. The department has anticipated possible law suits and harrassment by the Christian groups by making concessions in the wording of instructions to the performance assessments and in changing some of their plans to collect data. The Arizona department of education decided not to collect non-test indicators of learning from students or parents, and not to print a student profile that would aggregate achievement and related information on individual students for use in parent-teacher conferences. The Christian fundamentalists objected to collecting these data on the grounds of invasion of privacy, since the questions asked, for example, about how often and how much parents read to children and how much television the children watched in the home.

Chapter 1 reporting regulations loom as the major issue for state departments of education who want to move their statewide assessments to performance assessments. Most students in Chapter 1 programs are minority, immigrant, or speak little or no English. The regulations presently require a nationally aggregatable test, usually one of a list kept by the U.S. Department of Education. However, a number of states are talking with the USDE about accepting performance assessments for Chapter 1 reporting. Among them are Arizona and California. Arizona is making the case that Arizona Student Assessment Program gives teachers more information about the success of students than norm-referenced tests, where one or two point gains mean little or nothing.

It was fortuitous that we observed "We the People..." in Birmingham Alabama. (We could have observed it in any city where the competition was being held before Christmas 1990.) It was clear that the coordinators made every effort to include schools with African-American student populations. Where they had discretion to choose personnel (as in the case of the students who tallied the scores and some judges) the coordinators chose African-Americans. Nevertheless, it was distressing to see so few African-American students taking part in a competition devoted to the constitution in a city with such a profound connection with the Civil Rights movement.

Section VIII. Conclusion

We deliberately chose these six cases to show the variety of performance assessments, but the variety makes generalization hazardous. For example, it is difficult to say performance assessment is "new," when our cases include the Advanced Placement Studio Art Portfolio Assessment, which is 20 years old.

Nevertheless, two general lessons can be derived from these cases:

1. The purpose for which the assessment is intended should be clear before any decisions are made about form. Once made, the decision will dictate the components and the consequences of the assessment. This is a principle which applies not just to performance assessment but to all educational assessments: the purpose for which an assessment is intended may make a multiple-choice test the best candidate. At the same time, the original purpose can be ignored and the assessment used for unintended purposes, so the consequences should be carefully evaluated before choices are made about the form of the assessment.

2. The variety of performance assessments not only makes generalizations difficult, but it also introduces psychometric complications. Variety in educational assessment is both a blessing and a curse. While performance assessments can model instructional practice and assess progress in forms which do not distort the subject matter being assessed, they introduce unsolved problems of reliability in scoring. This is clear from more than half of our case-studies. In addition, equating scores across different kinds of assessments has been suggested as a means to measure educational progress across the nation. Our case studies should caution those who think this is easy or even possible.

A feature common to our case-studies is the involvement of a larger proportion of the educational community in performance assessment than the personnel usually involved in administering tests purchased from publishers. Even the Advanced Placement Studio Art assessment includes teachers among its readers (as do all Advanced Placement examinations), and two of the assessments include members of the community outside the school.

The ultimate aim of assessment should be to enable students to assess their own products. They can learn judgment best by seeing people around them exercising it, so the more teachers, administrators, parents, community members, and students themselves are involved, the better.

We do not want to underestimate the problems connected with performance assessment, (which will be clear from the six cases), but it should be emphasized that all are thoughtful approaches to directly assessing progress in learning. They have another important value: whether performance assessment replaces multiple-choice or not, these attempts to use it raise fundamental questions about what we value in education. The process of thinking about assessment is as valuable as the product.

SIX CASE-STUDIES OF PERFORMANCE ASSESSMENT

PREPARED FOR THE OFFICE OF TECHNOLOGY ASSESSMENT

BY

RUTH MITCHELL PH.D. AND AMY STEMPEL

COUNCIL FOR BASIC EDUCATION

Bibliography

Archbald, Doug A., and Fred M. Newmann. "The Nature of Authentic Academic Achievement." To be published in H. Berlak (ed.), Assessing Achievement: Toward the Development of a New Science of Educational Testing. Albany, NY: SUNY Press, forthcoming.

---------- Beyond Standardized Testing. Assessing Authentic Academic Achievement in the Secondary School. Reston, VA: National Association of Secondary School Principals (NASSP), 1988.

Askin, Walter. Evaluating the Advanced Placement Portfolio in Studio Art. Princeton, NJ: College Board and Educational Testing Service, 1986.

Barrs, Myra. "The Primary Language Record: Reflection of Issues in Evaluation." Language Arts, 67, 3 (March 1990), 244-253.

----------and Anne Thomas. The Primary Language Record Handbook. New York and London: Heinemann Educational Books, Inc., 1989.

----------. Patterns of Learning. London: Center for Language in Primary Education, 1990.

California Assessment Program. A Question of Thinking. A First Look at Students' Performance on Open-Ended Questions in Mathematics. Sacramento: California State Department of Education, 1989.

Cannell, John Jacob. Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average. West Virginia: Friends for Education, 1987

--------. How Public Educators Cheat on Standardized Achievement Tests. Albuquerque, NM:Friends for Education, 1989.

Commission on Excellence in Education. A Nation at Risk: The Imperative for Educational Reform. Washington DC: U.S. Government Printing Office, April 1983

De Lange, Jan. Mathematics, Insight, and Meaning. Utrecht, The Netherlands.

Dyer, Henry S. Parents Can Understand Testing. Columbia MD: National Committee for Citizens in Education, 1980.

Educational Leadership. Redirecting Assessment. 46: 7 (April 1989).

Educational Testing Service (ETS). Learning by Doing: A Manual for Teaching and Assessing Higher-Order Thinking in Science and Mathematics. NAEP Report # 17-HOS-80, May 1987.

Frederiksen, John R. and Allan Collins. "A Systems Approach to Educational Testing." Educational Researcher, 1989, 18 (9), 27-32.

Gardner, Howard. Frames of Mind. The Theory of Multiple Intelligences. New York: Basic Books, Inc., 1983

Gardner, Howard. "Assessment in Context: The Alternative to Standardized Testing." In B. Gifford and M.C. O'Connor, eds. Cognitive Approaches to Assessment. Boston: Kluwer Academic Publishers, forthcoming.

Gendler, Tamar. "Testing what we want to measure." Basic Education, 33,1 (September 1988), 7-11.

Goodman, Kenneth S.; Yetta M. Goodman, and Wendy J. Hood, eds. The Whole Language Evaluation Book. Portsmouth NH: Heinemann, and Toronto Canada: Irwin Publishing, 1989

Gould, Stephen Jay. The Mismeasure of Man. New York and London: W.W. Norton and company, 1981.

Houts, Paul L., ed. The Myth of Measurability. New York: Hart Publishing Company, 1977.

Howard, Kathryn. "Making the Writing Portfolio Real," The Quarterly of the National Writing Project and the Center for the Study of Writing, 12, 2 (Spring 1990), 4-7,27.

Kamii, Constance, ed. Achievement Testing in the Early Grades. Washington DC: National Association for the Education of Young Children (NAEYC), 1989.

King, Dorothy F. "Real Kids or Unreal Tasks: the Obvious Choice." Kentucky English Bulletin, Winter 1990.

Koretz, Daniel. "Arriving in Lake Wobegon. Are Standardized Tests Exaggerating Achievement and Distorting Instruction?" American Educator, 12, 2 (Summer 1988), 8-15

Kulm, Gerald, ed. Asssessing Higher Order Thinking in Mathematics. Washington DC: American Association for the Advancement of Science, 1990.

McKnight, Curtis C.; F. Joe Crosswhite; John A. Dossey; Edward Kifer; Jane O. Swafford; Kenneth J. Travers, and Thomas J. Cooney. The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective. Champaign, IL: Stipes Publishing Company, 1987.

Meier, Deborah. "Why reading tests don't test reading." Dissent, Fall 1981, pp.457-465.

------- "In education, small is sensible." New York Times, 8 September 1989, Op-Ed page.

------- and Ruth Jordan. "The right 'choice' for teachers." Teacher Magazine, December 1989, pp. 68-69.

Mitchell, Ruth. "Another failure for California Schools, where "A'stands for easy accreditation," Los Angeles Times, March 26, 1989, p. 3 and 6.

----------"Authentic Assessment." Basic Education, 33, 10, (June 1989), 6-10.

----------"Performance Assessment: An Emphasis on 'Activity.'" Education Week, 24 January 1990, p.56.

----------"A Letter to Test Publishers." Basic Education, 34, 7 (March 1990)??????

----------"The English National Assessment: Some First Hand Observations." Network News and Views, IX, 8 (August 1990), 79-85

----------"Deukmejian Gets Dunce Cap for Spiking Innovative School Assessment Program." Los Angeles Times, 19 August 1990, M5.

Mumme, Judith. Portfolio Assessment in Mathematics. Santa Barbara, CA: California Mathematics Project, 1990.

National Commission on Testing and Public Policy. From Gatekeeper to Gateway: Transforming Testing in America. Chestnut Hill, MA: National Commission on Testing and Public Policy, Boston College.

National Council of Teachers of English. Report Card on Basals.
Urbana, IL: NCTE, 1988.

National Council of Teachers of Mathematics (NCTM). Curriculum and Evaluation Standards for School Mathematics. MCTM, March 1989

National Research Council, Mathematical Sciences Education Board. Everybody Counts.
Washington, DC: National Academy Press, 1989.

Paulos, John Allen. I Think, Therefore I Laugh. New York: Columbia University Press, 1985.

---------- Innumeracy: Mathematical Illiteracy and Its Consequences. New York: Hill and Wang, 1988

Pearson. P. David and Sheila Valencia. "Assessment, Accountability, and Professional Prerogative," in Research in Literacy: Merging Perspectives. Thirty-Sixth Yearbook of The National Reading Conference. The National Reading Conference, Inc. 1987, 3-16.

Perrone, Vito. Working Papers: Reflections on Teachers, Schools, and Communities. New York and London: Teachers College Press, 1989

OERI State Accountability Study Group. Creating Responsible and Responsive Accountability systems. Washington DC: US Department of Education, September 1988.

Phi Delta Kappan. Special Section on Testing. 70:9 (May 1989).

Resnick, Lauren. Education and Learning to Think. Washington, DC: National Academy Press, 1987.

--------- and Leopold Klopfer, eds. Toward the thinking Curriculum: Current Cognitive Research. Reston, VA: The Association for Supervision and Curriculum Development, 1989.

Schwartz, Judah L. and Katherine A. Viator, eds. The Prices of Secrecy: The Social, Intellectual, and Psychological Costs of Current Assessment Practice. Cambridge, MA: Educational Technology Center (ETC), Harvard Graduate School of Education, September 1990.

Shanker, Albert. "The End of the Traditional Model of Schooling--and a Proposal for Using Incentives to Restructure Our Public Schools." Phi Delta Kappan, 71, 5 (January 1990), 345-357.

Steen, Lynn Arthur. "Teaching Mathematics for Tomorrow's World." Educational Leadership, 47,1 (September 1989), 18-23.

Suen, Hoi K. and Bruce Davey. "Potential theoretical and practical pitfalls and cautions of the performance assessment design." Paper presented at the 1990 Annual Meeting of the American Educational Research Association (AERA).

The OSS Assessment Staff. Assessment of Men. Selection of Personnel for the Office of Strategic Services. New York: Rinehart and Company, 1948.

Vermont Department of Education. Should These Be Vermont's Goals for Education? Montpelier VT: Department of Education, 1989.

Washington Post. "Beyond the No.2 Pencil." Editorial page, 28 November 1989.

Wiggins, Grant. "Creating a Thought Provoking Curriculum." American Educator. Winter 1987.